Elements of
Abstract and Linear Algebra

E. H. Connell

E.H. Connell
Department of Mathematics
University of Miami
P.O. Box 249085
Coral Gables, Florida 33124 USA
ec@math.miami.edu

Mathematical Subject Classifications (1991): 12-01, 13-01, 15-01, 16-01, 20-01

March 20, 2004

Introduction

In 1965 I first taught an undergraduate course in abstract algebra. It was fun to teach because the material was interesting and the class was outstanding. Five of those students later earned a Ph.D. in mathematics. Since then I have taught the course about a dozen times from various texts. Over the years I developed a set of lecture notes and in 1985 I had them typed so they could be used as a text. They now appear (in modified form) as the first five chapters of this book. Here were some of my motives at the time.

1) To have something as short and inexpensive as possible. In my experience, students like short books.

2) To avoid all innovation. To organize the material in the most simple-minded straightforward manner.

3) To order the material linearly. To the extent possible, each section should use the previous sections and be used in the following sections.

4) To omit as many topics as possible. This is a foundational course, not a topics course. If a topic is not used later, it should not be included. There are three good reasons for this. First, linear algebra has top priority. It is better to go forward and do more linear algebra than to stop and do more group and ring theory. Second, it is more important that students learn to organize and write proofs themselves than to cover more subject matter. Algebra is a perfect place to get started because there are many "easy" theorems to prove. There are many routine theorems stated here without proofs, and they may be considered as exercises for the students. Third, the material should be so fundamental that it be appropriate for students in the physical sciences and in computer science. Zillions of students take calculus and cookbook linear algebra, but few take abstract algebra courses. Something is wrong here, and one thing wrong is that the courses try to do too much group and ring theory and not enough matrix theory and linear algebra.

5) To offer an alternative for computer science majors to the standard discrete mathematics courses. Most of the material in the first four chapters of this text is covered in various discrete mathematics courses. Computer science majors might benefit by seeing this material organized from a purely mathematical viewpoint.

Over the years I used the five chapters that were typed as a base for my algebra courses, supplementing them as I saw fit. In 1996 I wrote a sixth chapter, giving enough material for a full first year graduate course. This chapter was written in the same "style" as the previous chapters, i.e., everything was right down to the nub. It hung together pretty well except for the last two sections on determinants and dual spaces. These were independent topics stuck on at the end. In the academic year 1997-98 I revised all six chapters and had them typed in LaTeX. This is the personal background of how this book came about.

It is difficult to do anything in life without help from friends, and many of my friends have contributed to this text. My sincere gratitude goes especially to Marilyn Gonzalez, Lourdes Robles, Marta Alpar, John Zweibel, Dmitry Gokhman, Brian Coomes, Huseyin Kocak, and Shulim Kaliman. To these and all who contributed, this book is fondly dedicated.

This book is a survey of abstract algebra with emphasis on linear algebra. It is intended for students in mathematics, computer science, and the physical sciences. The first three or four chapters can stand alone as a one semester course in abstract algebra. However they are structured to provide the background for the chapter on linear algebra. Chapter 2 is the most difficult part of the book because groups are written in additive and multiplicative notation, and the concept of coset is confusing at first. After Chapter 2 the book gets easier as you go along. Indeed, after the first four chapters, the linear algebra follows easily. Finishing the chapter on linear algebra gives a basic one year undergraduate course in abstract algebra. Chapter 6 continues the material to complete a first year graduate course. Classes with little background can do the first three chapters in the first semester, and chapters 4 and 5 in the second semester. More advanced classes can do four chapters the first semester and chapters 5 and 6 the second semester. As bare as the first four chapters are, you still have to truck right along to finish them in one semester.

The presentation is compact and tightly organized, but still somewhat informal. The proofs of many of the elementary theorems are omitted. These proofs are to be provided by the professor in class or assigned as homework exercises. There is a non-trivial theorem stated without proof in Chapter 4, namely the determinant of the product is the product of the determinants. For the proper flow of the course, this theorem should be assumed there without proof. The proof is contained in Chapter 6. The Jordan form should not be considered part of Chapter 5. It is stated there only as a reference for undergraduate courses. Finally, Chapter 6 is not written primarily for reference, but as an additional chapter for more advanced courses.

This text is written with the conviction that it is more effective to teach abstract and linear algebra as one coherent discipline rather than as two separate ones. Teaching abstract algebra and linear algebra as distinct courses results in a loss of synergy and a loss of momentum. Also with this text the professor does not extract the course from the text, but rather builds the course upon it. I am convinced it is easier to build a course from a base than to extract it from a big book. Because after you extract it, you still have to build it. The bare bones nature of this book adds to its flexibility, because you can build whatever course you want around it. Basic algebra is a subject of incredible elegance and utility, but it requires a lot of organization. This book is my attempt at that organization. Every effort has been extended to make the subject move rapidly and to make the flow from one topic to the next as seamless as possible. The student has limited time during the semester for serious study, and this time should be allocated with care. The professor picks which topics to assign for serious study and which ones to "wave arms at". The goal is to stay focused and go forward, because mathematics is learned in hindsight. I would have made the book shorter, but I did not have any more time.

When using this text, the student already has the outline of the next lecture, and each assignment should include the study of the next few pages. Study forward, not just back. A few minutes of preparation does wonders to leverage classroom learning, and this book is intended to be used in that manner. The purpose of class is to learn, not to do transcription work. When students come to class cold and spend the period taking notes, they participate little and learn little. This leads to a dead class and also to the bad psychology of "O K, I am here, so teach me the subject." Mathematics is not taught, it is learned, and many students never learn how to learn. Professors should give more direction in that regard.

Unfortunately mathematics is a difficult and heavy subject. The style and approach of this book is to make it a little lighter. This book works best when viewed lightly and read as a story. I hope the students and professors who try it, enjoy it.

E. H. Connell

Department of Mathematics
University of Miami
Coral Gables, FL 33124
ec@math.miami.edu

Outline

Chapter 1 Background and Fundamentals of Mathematics

Chapter 2 Groups

Chapter 3 Rings

Chapter 4 Matrices and Matrix Rings

1	2	3	4
5	6	7	8
9	11	10	

Abstract algebra is not only a major subject of science, but it is also magic and fun. Abstract algebra is not all work and no play, and it is certainly not a dull boy. See, for example, the neat card trick on page 18. This trick is based, not on sleight of hand, but rather on a theorem in abstract algebra. Anyone can do it, but to understand it you need some group theory. And before beginning the course, you might first try your skills on the famous (some would say infamous) tile puzzle. In this puzzle, a frame has 12 spaces, the first 11 with numbered tiles and the last vacant. The last two tiles are out of order. Is it possible to slide the tiles around to get them all in order, and end again with the last space vacant? After giving up on this, you can study permutation groups and learn the answer!

Chapter 1

Background and Fundamentals of Mathematics

This chapter is fundamental, not just for algebra, but for all fields related to mathematics. The basic concepts are products of sets, partial orderings, equivalence relations, functions, and the integers. An equivalence relation on a set A is shown to be simply a partition of A into disjoint subsets. There is an emphasis on the concept of function, and the properties of surjective, injective, and bijective. The notion of a solution of an equation is central in mathematics, and most properties of functions can be stated in terms of solutions of equations. In elementary courses the section on the Hausdorff Maximality Principle should be ignored. The final section gives a proof of the unique factorization theorem for the integers.

Notation Mathematics has its own universally accepted shorthand. The symbol $\exists$ means "there exists" and $\exists!$ means "there exists a unique". The symbol $\forall$ means "for each" and $\Rightarrow$ means "implies". Some sets (or collections) are so basic they have their own proprietary symbols. Five of these are listed below.

$\mathbf{N} = \mathbf{Z}^+ =$ the set of positive integers $= \{1, 2, 3, ...\}$
$\mathbf{Z}$ = the ring of integers $= \{..., -2, -1, 0, 1, 2, ...\}$
$\mathbf{Q}$ = the field of rational numbers $= \{a/b : a, b \in \mathbf{Z}, b \neq 0\}$
$\mathbf{R}$ = the field of real numbers
$\mathbf{C}$ = the field of complex numbers $= \{a + bi : a, b \in \mathbf{R}\}$ $(i^2 = -1)$

Sets Suppose $A, B, C,...$ are sets. We use the standard notation for intersection and union.

$A \cap B = \{x : x \in A \text{ and } x \in B\}$ = the set of all x which are elements

of A and B.

$A \cup B = \{x : x \in A \text{ or } x \in B\}$ = the set of all x which are elements of A or B.

Any set called an index set is assumed to be non-void. Suppose T is an index set and for each $t \in T$, A_t is a set.

$$\bigcup_{t \in T} A_t = \{x : \exists\ t \in T \text{ with } x \in A_t\}$$

$$\bigcap_{t \in T} A_t = \{x : \text{if } t \in T, x \in A_t\} = \{x : \forall t \in T, x \in A_t\}$$

Let $\emptyset$ be the null set. If $A \cap B = \emptyset$, then A and B are said to be *disjoint*.

Definition Suppose each of A and B is a set. The statement that A is a subset of B $(A \subset B)$ means that if a is an element of A, then a is an element of B. That is, $a \in A \Rightarrow a \in B$. If $A \subset B$ we may say A is contained in B, or B contains A.

Exercise Suppose each of A and B is a set. The statement that A is not a subset of B means __________________________.

Theorem (De Morgan's laws) Suppose S is a set. If $C \subset S$ (i.e., if C is a subset of S), let C', the complement of C in S, be defined by $C' = S - C = \{x \in S : x \notin C\}$. Then for any $A, B \subset S$,

$$(A \cap B)' = A' \cup B' \quad \text{and}$$

$$(A \cup B)' = A' \cap B'$$

Cartesian Products If X and Y are sets, $X \times Y = \{(x, y) : x \in X \text{ and } y \in Y\}$. In other words, the Cartesian product of X and Y is defined to be the set of all ordered pairs whose first term is in X and whose second term is in Y.

Example $\mathbf{R} \times \mathbf{R} = \mathbf{R}^2$ = the plane.

Definition If each of $X_1, ..., X_n$ is a set, $X_1 \times \cdots \times X_n = \{(x_1, ..., x_n) : x_i \in X_i$ for $1 \le i \le n\}$ = the set of all ordered n-tuples whose i-th term is in X_i.

Example $\mathbf{R} \times \cdots \times \mathbf{R} = \mathbf{R}^n$ = real n-space.

Question Is $(\mathbf{R} \times \mathbf{R}^2) = (\mathbf{R^2} \times \mathbf{R}) = \mathbf{R^3}$?

Relations

If A is a non-void set, a non-void subset $R \subset A \times A$ is called a *relation* on A. If $(a, b) \in R$ we say that a is related to b, and we write this fact by the expression $a \sim b$. Here are several properties which a relation may possess.

1) If $a \in A$, then $a \sim a$. (reflexive)
2) If $a \sim b$, then $b \sim a$. (symmetric)
2′) If $a \sim b$ and $b \sim a$, then $a = b$. (anti-symmetric)
3) If $a \sim b$ and $b \sim c$, then $a \sim c$. (transitive)

Definition A relation which satisfies 1), 2′), and 3) is called a *partial ordering.* In this case we write $a \sim b$ as $a \le b$. Then

1) If $a \in A$, then $a \le a$.
2′) If $a \le b$ and $b \le a$, then $a = b$.
3) If $a \le b$ and $b \le c$, then $a \le c$.

Definition A *linear ordering* is a partial ordering with the additional property that, if $a, b \in A$, then $a \le b$ or $b \le a$.

Example $A = \mathbf{R}$ with the ordinary ordering, is a linear ordering.

Example A = all subsets of $\mathbf{R}^2$, with $a \le b$ defined by $a \subset b$, is a partial ordering.

Hausdorff Maximality Principle (HMP) Suppose S is a non-void subset of A and $\sim$ is a relation on A. This defines a relation on S. If the relation satisfies any of the properties 1), 2), 2′), or 3) on A, the relation also satisfies these properties when restricted to S. In particular, a partial ordering on A defines a partial ordering

on S. However the ordering may be linear on S but not linear on A. The HMP is that any linearly ordered subset of a partially ordered set is contained in a maximal linearly ordered subset.

Exercise Define a relation on $A = \mathbf{R}^2$ by $(a, b) \sim (c, d)$ provided $a \leq c$ and $b \leq d$. Show this is a partial ordering which is linear on $S = \{(a, a) : a < 0\}$. Find at least two maximal linearly ordered subsets of $\mathbf{R}^2$ which contain S.

One of the most useful applications of the HMP is to obtain maximal monotonic collections of subsets.

Definition A collection of sets is said to be *monotonic* if, given any two sets of the collection, one is contained in the other.

Corollary to HMP Suppose X is a non-void set and A is some non-void collection of subsets of X, and S is a subcollection of A which is monotonic. Then $\exists$ a maximal monotonic subcollection of A which contains S.

Proof Define a partial ordering on A by $V \leq W$ iff $V \subset W$, and apply HMP.

The HMP is used twice in this book. First, to show that infinitely generated vector spaces have free bases, and second, in the Appendix, to show that rings have maximal ideals (see pages 87 and 109). In each of these applications, the maximal monotonic subcollection will have a maximal element. In elementary courses, these results may be assumed, and thus the HMP may be ignored.

Equivalence Relations A relation satisfying properties 1), 2), and 3) is called an *equivalence relation*.

Exercise Define a relation on $A = \mathbf{Z}$ by $n \sim m$ iff $n - m$ is a multiple of 3. Show this is an equivalence relation.

Definition If $\sim$ is an equivalence relation on A and $a \in A$, we define the *equivalence class* containing a by $cl(a) = \{x \in A : a \sim x\}$.

Theorem

1) If $b \in cl(a)$ then $cl(b) = cl(a)$. Thus we may speak of a subset of A being an equivalence class with no mention of any element contained in it.
2) If each of $U, V \subset A$ is an equivalence class and $U \cap V \neq \emptyset$, then $U = V$.
3) Each element of A is an element of one and only one equivalence class.

Definition A *partition* of A is a collection of disjoint non-void subsets whose union is A. In other words, a collection of non-void subsets of A is a partition of A provided any $a \in A$ is an element of one and only one subset of the collection. Note that if A has an equivalence relation, the equivalence classes form a partition of A.

Theorem Suppose A is a non-void set with a partition. Define a relation on A by $a \sim b$ iff a and b belong to the same subset of the partition. Then $\sim$ is an equivalence relation, and the equivalence classes are just the subsets of the partition.

Summary There are two ways of viewing an equivalence relation — one is as a relation on A satisfying 1), 2), and 3), and the other is as a partition of A into disjoint subsets.

Exercise Define an equivalence relation on $\mathbf{Z}$ by $n \sim m$ iff $n - m$ is a multiple of 3. What are the equivalence classes?

Exercise Is there a relation on $\mathbf{R}$ satisfying 1), 2), 2′) and 3) ? That is, is there an equivalence relation on $\mathbf{R}$ which is also a partial ordering?

Exercise Let $H \subset \mathbf{R}^2$ be the line $H = \{(a, 2a) : a \in \mathbf{R}\}$. Consider the collection of all translates of H, i.e., all lines in the plane with slope 2. Find the equivalence relation on $\mathbf{R}^2$ defined by this partition of $\mathbf{R}^2$.

Functions

Just as there are two ways of viewing an equivalence relation, there are two ways of defining a function. One is the "intuitive" definition, and the other is the "graph" or "ordered pairs" definition. In either case, *domain* and *range* are inherent parts of the definition. We use the "intuitive" definition because everyone thinks that way.

Definition If X and Y are (non-void) sets, a *function* or *mapping* or *map* with domain X and range Y, is an ordered triple (X, Y, f) where f assigns to each $x \in X$ a well defined element $f(x) \in Y$. The statement that (X, Y, f) is a function is written as $f : X \to Y$ or $X \xrightarrow{f} Y$.

Definition The *graph* of a function (X, Y, f) is the subset $\Gamma \subset X \times Y$ defined by $\Gamma = \{(x, f(x)) : x \in X\}$. The connection between the "intuitive" and "graph" viewpoints is given in the next theorem.

Theorem If $f : X \to Y$, then the graph $\Gamma \subset X \times Y$ has the property that each $x \in X$ is the first term of one and only one ordered pair in Γ. Conversely, if Γ is a subset of $X \times Y$ with the property that each $x \in X$ is the first term of one and only ordered pair in Γ, then $\exists !\ f : X \to Y$ whose graph is Γ. The function is defined by "$f(x)$ is the second term of the ordered pair in Γ whose first term is x."

Example *Identity functions* Here $X = Y$ and $f : X \to X$ is defined by $f(x) = x$ for all $x \in X$. The identity on X is denoted by I_X or just $I : X \to X$.

Example *Constant functions* Suppose $y_0 \in Y$. Define $f : X \to Y$ by $f(x) = y_0$ for all $x \in X$.

Restriction Given $f : X \to Y$ and a non-void subset S of X, define $f \mid S : S \to Y$ by $(f \mid S)(s) = f(s)$ for all $s \in S$.

Inclusion If S is a non-void subset of X, define the inclusion $i : S \to X$ by $i(s) = s$ for all $s \in S$. Note that inclusion is a restriction of the identity.

Composition Given $W \xrightarrow{f} X \xrightarrow{g} Y$ define $g \circ f : W \to Y$ by $(g \circ f)(x) = g(f(x))$.

Theorem (The associative law of composition) If $V \xrightarrow{f} W \xrightarrow{g} X \xrightarrow{h} Y$, then $h \circ (g \circ f) = (h \circ g) \circ f$. This may be written as $h \circ g \circ f$.

Definitions Suppose $f : X \to Y$.

1) If $T \subset Y$, the *inverse image of* T is a subset of X, $f^{-1}(T) = \{x \in X : f(x) \in T\}$.

2) If $S \subset X$, the *image of* S is a subset of Y, $f(S) = \{f(s) : s \in S\} = \{y \in Y : \exists s \in S \text{ with } f(s) = y\}$.

3) The *image of* f is the image of X , i.e., image $(f) = f(X) = \{f(x) : x \in X\} = \{y \in Y : \exists x \in X \text{ with } f(x) = y\}$.

4) $f : X \to Y$ is *surjective* or *onto* provided image $(f) = Y$ i.e., the image is the range, i.e., if $y \in Y$, $f^{-1}(y)$ is a non-void subset of X.

5) $f : X \to Y$ is *injective* or *1-1* provided $(x_1 \neq x_2) \Rightarrow f(x_1) \neq f(x_2)$, i.e., if x_1 and x_2 are distinct elements of X, then $f(x_1)$ and $f(x_2)$ are distinct elements of Y.

6) $f : X \to Y$ is *bijective* or is a *1-1 correspondence* provided f is surjective and injective. In this case, there is function $f^{-1} : Y \to X$ with $f^{-1} \circ f = I_X : X \to X$ and $f \circ f^{-1} = I_Y : Y \to Y$. Note that $f^{-1} : Y \to X$ is also bijective and $(f^{-1})^{-1} = f$.

Examples

1) $f : \mathbf{R} \to \mathbf{R}$ defined by $f(x) = \sin(x)$ is neither surjective nor injective.

2) $f : \mathbf{R} \to [-1, 1]$ defined by $f(x) = \sin(x)$ is surjective but not injective.

3) $f : [0, \pi/2] \to \mathbf{R}$ defined by $f(x) = \sin(x)$ is injective but not surjective.

4) $f : [0, \pi/2] \to [0, 1]$ defined by $f(x) = \sin(x)$ is bijective. ($f^{-1}(x)$ is written as $\arcsin(x)$ or $\sin^{-1}(x)$.)

5) $f : \mathbf{R} \to (0, \infty)$ defined by $f(x) = e^x$ is bijective. ($f^{-1}(x)$ is written as $\ln(x)$.)

Note There is no such thing as "the function $\sin(x)$." A function is not defined unless the domain and range are specified.

Exercise Show there are natural bijections from $(\mathbf{R} \times \mathbf{R}^2)$ to $(\mathbf{R}^2 \times \mathbf{R})$ and from $(\mathbf{R}^2 \times \mathbf{R})$ to $\mathbf{R} \times \mathbf{R} \times \mathbf{R}$. These three sets are disjoint, but the bijections between them are so natural that we sometimes identify them.

Exercise Suppose X is a set with 6 elements and Y is a finite set with n elements.

1) There exists an injective $f : X \to Y$ iff n ________________.
2) There exists a surjective $f : X \to Y$ iff n ________________.
3) There exists a bijective $f : X \to Y$ iff n ________________.

Pigeonhole Principle Suppose X is a finite set with m elements, Y is a finite set with n elements, and $f : X \to Y$ is a function.

1) If $m = n$, then f is injective iff f is surjective iff f is bijective.
2) If $m > n$, then f is not injective.
3) If $m < n$, then f is not surjective.

If you are placing 6 pigeons in 6 holes, and you run out of pigeons before you fill the holes, then you have placed 2 pigeons in one hole. In other words, in part 1) for $m = n = 6$, if f is not surjective then f is not injective. Of course, the pigeonhole principle does not hold for infinite sets, as can be seen by the following exercise.

Exercise Show there is a function $f : \mathbf{Z}^+ \to \mathbf{Z}^+$ which is injective but not surjective. Also show there is one which is surjective but not injective.

Exercise Suppose $f : [-2, 2] \to \mathbf{R}$ is defined by $f(x) = x^2$. Find $f^{-1}(f([1, 2]))$. Also find $f(f^{-1}([3, 5]))$.

Exercise Suppose $f : X \to Y$ is a function, $S \subset X$ and $T \subset Y$. Find the relationship between S and $f^{-1}(f(S))$. Show that if f is injective, $S = f^{-1}(f(S))$. Also find the relationship between T and $f(f^{-1}(T))$. Show that if f is surjective, $T = f(f^{-1}(T))$.

Strips We now define the vertical and horizontal strips of $X \times Y$.

If $x_0 \in X$, $\{(x_0, y) : y \in Y\} = (x_0 \times Y)$ is called a *vertical strip.*
If $y_0 \in Y$, $\{(x, y_0) : x \in X\} = (X \times y_0)$ is called a *horizontal strip.*

Theorem Suppose $S \subset X \times Y$. The subset S is the graph of a function with domain X and range Y iff each vertical strip intersects S in exactly one point.

This is just a restatement of the property of a graph of a function. The purpose of the next theorem is to restate properties of functions in terms of horizontal strips.

Theorem Suppose $f : X \to Y$ has graph Γ. Then

1) Each horizontal strip intersects Γ in at least one point iff f is ________.
2) Each horizontal strip intersects Γ in at most one point iff f is ________.
3) Each horizontal strip intersects Γ in exactly one point iff f is ________.

Solutions of Equations Now we restate these properties in terms of solutions of equations. Suppose $f : X \to Y$ and $y_0 \in Y$. Consider the equation $f(x) = y_0$. Here y_0 is given and x is considered to be a "variable". A *solution* to this equation is any $x_0 \in X$ with $f(x_0) = y_0$. Note that the set of all solutions to $f(x) = y_0$ is $f^{-1}(y_0)$. Also $f(x) = y_0$ has a solution iff $y_0 \in \text{image}(f)$ iff $f^{-1}(y_0)$ is non-void.

Theorem Suppose $f : X \to Y$.

1) The equation $f(x) = y_0$ has at least one solution for each $y_0 \in Y$ iff f is ____________________.
2) The equation $f(x) = y_0$ has at most one solution for each $y_0 \in Y$ iff f is ____________________.
3) The equation $f(x) = y_0$ has a unique solution for each $y_0 \in Y$ iff f is ____________________.

Right and Left Inverses One way to understand functions is to study right and left inverses, which are defined after the next theorem.

Theorem Suppose $X \xrightarrow{f} Y \xrightarrow{g} W$ are functions.

1) If $g \circ f$ is injective, then f is injective.

2) If $g \circ f$ is surjective, then g is surjective.
3) If $g \circ f$ is bijective, then f is injective and g is surjective.

Example $X = W = \{p\}$, $Y = \{p, q\}$, $f(p) = p$, and $g(p) = g(q) = p$. Here $g \circ f$ is the identity, but f is not surjective and g is not injective.

Definition Suppose $f : X \to Y$ is a function. A left inverse of f is a function $g : Y \to X$ such that $g \circ f = I_X : X \to X$. A right inverse of f is a function $h : Y \to X$ such that $f \circ h = I_Y : Y \to Y$.

Theorem Suppose $f : X \to Y$ is a function.

1) f has a right inverse iff f is surjective. Any such right inverse must be injective.
2) f has a left inverse iff f is injective. Any such left inverse must be surjective.

Corollary Suppose each of X and Y is a non-void set. Then $\exists$ an injective $f : X \to Y$ iff $\exists$ a surjective $g : Y \to X$. Also a function from X to Y is bijective iff it has a left inverse and a right inverse iff it has a left and right inverse.

Note The Axiom of Choice is not discussed in this book. However, if you worked 1) of the theorem above, you unknowingly used one version of it. For completeness, we state this part of 1) again.

The Axiom of Choice If $f : X \to Y$ is surjective, then f has a right inverse h. That is, for each $y \in Y$, it is possible to choose an $x \in f^{-1}(y)$ and thus to define $h(y) = x$.

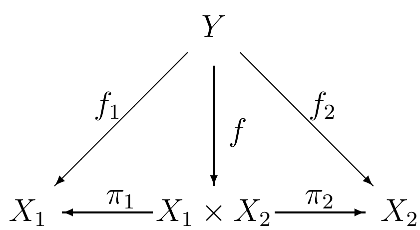
Note It is a classical theorem in set theory that the Axiom of Choice and the Hausdorff Maximality Principle are equivalent. However in this text we do not go that deeply into set theory. For our purposes it is assumed that the Axiom of Choice and the HMP are true.

Exercise Suppose $f : X \to Y$ is a function. Define a relation on X by $a \sim b$ if $f(a) = f(b)$. Show this is an equivalence relation. If y belongs to the image of f, then $f^{-1}(y)$ is an equivalence class and every equivalence class is of this form. In the next chapter where f is a group homomorphism, these equivalence classes will be called cosets.

Projections If X_1 and X_2 are non-void sets, we define the projection maps $\pi_1 : X_1 \times X_2 \to X_1$ and $\pi_2 : X_1 \times X_2 \to X_2$ by $\pi_i(x_1, x_2) = x_i$.

Theorem If Y, X_1, and X_2 are non-void sets, there is a 1-1 correspondence between $\{$functions $f \colon Y \to X_1 \times X_2\}$ and $\{$ordered pairs of functions (f_1, f_2) where $f_1 \colon Y \to X_1$ and $f_2 : Y \to X_2\}$.

Proof Given f, define $f_1 = \pi_1 \circ f$ and $f_2 = \pi_2 \circ f$. Given f_1 and f_2 define $f : Y \to X_1 \times X_2$ by $f(y) = (f_1(y), f_2(y))$. Thus a function from Y to $X_1 \times X_2$ is merely a pair of functions from Y to X_1 and Y to X_2. This concept is displayed in the diagram below. It is summarized by the equation $f = (f_1, f_2)$.



One nice thing about this concept is that it works fine for infinite Cartesian products.

Definition Suppose T is an index set and for each $t \in T$, X_t is a non-void set. Then the *product* $\prod\limits_{t \in T} X_t = \prod X_t$ is the collection of all sequences $\{x_t\}_{t \in T} = \{x_t\}$ where $x_t \in X_t$. Formally these sequences are functions α from T to $\bigcup X_t$ with each $\alpha(t)$ in X_t and written as $\alpha(t) = x_t$. If $T = \{1, 2, \ldots, n\}$ then $\{x_t\}$ is the ordered n-tuple $(x_1, x_2, \ldots, x_n)$. If $T = \mathbf{Z}^+$ then $\{x_t\}$ is the sequence $(x_1, x_2, \ldots)$. For any T and any s in T, the *projection map* $\pi_s : \prod X_t \to X_s$ is defined by $\pi_s(\{x_t\}) = x_s$.

Theorem If Y is any non-void set, there is a 1-1 correspondence between $\{$functions $f : Y \to \prod X_t\}$ and $\{$sequences of functions $\{f_t\}_{t \in T}$ where $f_t : Y \to X_t\}$. Given f, the sequence $\{f_t\}$ is defined by $f_t = \pi_t \circ f$. Given $\{f_t\}$, f is defined by $f(y) = \{f_t(y)\}$.

A Calculus Exercise Let A be the collection of all functions $f : [0,1] \rightarrow \mathbf{R}$ which have an infinite number of derivatives. Let $A_0 \subset A$ be the subcollection of those functions f with $f(0) = 0$. Define $D : A_0 \rightarrow A$ by $D(f) = df/dx$. Use the mean value theorem to show that D is injective. Use the fundamental theorem of calculus to show that D is surjective.

Exercise This exercise is not used elsewhere in this text and may be omitted. It is included here for students who wish to do a little more set theory. Suppose T is a non-void set.

1) If Y is a non-void set, define Y^T to be the collection of all functions with domain T and range Y. Show that if T and Y are finite sets with m and n elements, then Y^T has n^m elements. In particular, when $T = \{1,2,3\}$, $Y^T = Y \times Y \times Y$ has n^3 elements. Show that if $n \geq 3$, the subset of $Y^{\{1,2,3\}}$ of all injective functions has $n(n-1)(n-2)$ elements. These injective functions are called permutations on Y taken 3 at a time. If $T = \mathbf{N}$, then Y^T is the infinite product $Y \times Y \times \cdots$. That is, $Y^{\mathbf{N}}$ is the set of all infinite sequences $(y_1, y_2, \ldots)$ where each $y_i \in Y$. For any Y and T, let Y_t be a copy of Y for each $t \in T$. Then $Y^T = \prod_{t \in T} Y_t$.

2) Suppose each of Y_1 and Y_2 is a non-void set. Show there is a natural bijection from $(Y_1 \times Y_2)^T$ to $Y_1^T \times Y_2^T$. (This is the fundamental property of Cartesian products presented in the two previous theorems.)

3) Define $\mathcal{P}(T)$, the power set of T, to be the collection of all subsets of T (including the null set). Show that if T is a finite set with m elements, $\mathcal{P}(T)$ has 2^m elements.

4) If S is any subset of T, define its characteristic function $\chi_S : T \rightarrow \{0,1\}$ by letting $\chi_S(t)$ be 1 when $t \in S$, and be 0 when $t \notin S$. Define $\alpha : \mathcal{P}(T) \rightarrow \{0,1\}^T$ by $\alpha(S) = \chi_S$. Define $\beta : \{0,1\}^T \rightarrow \mathcal{P}(T)$ by $\beta(f) = f^{-1}(1)$. Show that if $S \subset T$ then $\beta \circ \alpha(S) = S$, and if $f : T \rightarrow \{0,1\}$ then $\alpha \circ \beta(f) = f$. Thus α is a bijection and $\beta = \alpha^{-1}$.

$$\mathcal{P}(T) \longleftrightarrow \{0,1\}^T$$

5) Suppose $\gamma : T \rightarrow \{0,1\}^T$ is a function and show that it cannot be surjective. If $t \in T$, denote $\gamma(t)$ by $\gamma(t) = f_t : T \rightarrow \{0,1\}$. Define $f : T \rightarrow \{0,1\}$ by $f(t) = 0$ if $f_t(t) = 1$, and $f(t) = 1$ if $f_t(t) = 0$. Show that f is not in the image of γ and thus γ cannot be surjective. This shows that if T is an infinite set, then the set $\{0,1\}^T$ represents a "higher order of infinity than T".

6) An infinite set Y is said to be *countable* if there is a bijection from the positive

integers $\mathbf{N}$ to Y. Show $\mathbf{Q}$ is countable but the following three collections are not.

i) $\mathcal{P}(\mathbf{N})$, the collection of all subsets of $\mathbf{N}$.
ii) $\{0,1\}^{\mathbf{N}}$, the collection of all functions $f : \mathbf{N} \to \{0,1\}$.
iii) The collection of all sequences $(y_1, y_2, \ldots)$ where each y_i is 0 or 1.

We know that ii) and iii) are equal and there is a natural bijection between i) and ii). We also know there is no surjective map from $\mathbf{N}$ to $\{0,1\}^{\mathbf{N}}$, i.e., $\{0,1\}^{\mathbf{N}}$ is uncountable. Finally, show there is a bijection from $\{0,1\}^{\mathbf{N}}$ to the real numbers $\mathbf{R}$. (This is not so easy. To start with, you have to decide what the real numbers are.)

Notation for the Logic of Mathematics

Each of the words "Lemma", "Theorem", and "Corollary" means "true statement". Suppose A and B are statements. A theorem may be stated in any of the following ways:

Theorem **Hypothesis** Statement A.
Conclusion Statement B.

Theorem Suppose A is true. Then B is true.

Theorem If A is true, then B is true.

Theorem $A \Rightarrow B$ (A implies B).

There are two ways to prove the theorem — to suppose A is true and show B is true, or to suppose B is false and show A is false. The expressions "$A \Leftrightarrow B$", "A is equivalent to B", and "A is true iff B is true " have the same meaning (namely, that $A \Rightarrow B$ and $B \Rightarrow A$).

The important thing to remember is that thoughts and expressions flow through the language. Mathematical symbols are shorthand for phrases and sentences in the English language. For example, "$x \in B$ " means "x is an element of the set B." If A is the statement "$x \in \mathbf{Z}^+$" and B is the statement "$x^2 \in \mathbf{Z}^+$", then "$A \Rightarrow B$"means "If x is a positive integer, then x^2 is a positive integer".

Mathematical Induction is based upon the fact that if $S \subset \mathbf{Z}^+$ is a non-void subset, then S contains a smallest element.

Theorem Suppose $P(n)$ is a statement for each $n = 1, 2, ...$. Suppose $P(1)$ is true and for each $n \geq 1$, $P(n) \Rightarrow P(n+1)$. Then for each $n \geq 1$, $P(n)$ is true.

Proof If the theorem is false, then $\exists$ a smallest positive integer m such that $P(m)$ is false. Since $P(m-1)$ is true, this is impossible.

Exercise Use induction to show that, for each $n \geq 1$, $1+2+\cdots+n = n(n+1)/2$.

The Integers

In this section, lower case letters $a, b, c, ...$ will represent integers, i.e., elements of $\mathbf{Z}$. Here we will establish the following three basic properties of the integers.

1) If G is a subgroup of $\mathbf{Z}$, then $\exists\ n \geq 0$ such that $G = n\mathbf{Z}$.
2) If a and b are integers, not both zero, and G is the collection of all linear combinations of a and b, then G is a subgroup of $\mathbf{Z}$, and its positive generator is the greatest common divisor of a and b.
3) If $n \geq 2$, then n factors uniquely as the product of primes.

All of this will follow from long division, which we now state formally.

Euclidean Algorithm Given a, b with $b \neq 0$, $\exists!$ m and r with $0 \leq r < |b|$ and $a = bm + r$. In other words, b divides a "m times with a remainder of r". For example, if $a = -17$ and $b = 5$, then $m = -4$ and $r = 3$, $-17 = 5(-4) + 3$.

Definition If $r = 0$, we say that b *divides* a or a is a *multiple* of b. This fact is written as $b \mid a$. Note that $b \mid a \Leftrightarrow$ the rational number a/b is an integer $\Leftrightarrow \exists!\ m$ such that $a = bm \Leftrightarrow a \in b\mathbf{Z}$.

Note Anything (except 0) divides 0. 0 does not divide anything. $\pm$ 1 divides anything . If $n \neq 0$, the set of integers which n divides is $n\mathbf{Z} = \{nm : m \in \mathbf{Z}\} = \{..., -2n, -n, 0, n, 2n, ...\}$. Also n divides a and b with the same remainder iff n divides $(a - b)$.

Definition A non-void subset $G \subset \mathbf{Z}$ is a *subgroup* provided $(g \in G \Rightarrow -g \in G)$ and $(g_1, g_2 \in G \Rightarrow (g_1 + g_2) \in G)$. We say that G is closed under negation and closed under addition.

Theorem If $n \in \mathbf{Z}$ then $n\mathbf{Z}$ is a subgroup. Thus if $n \neq 0$, the set of integers which n divides is a subgroup of $\mathbf{Z}$.

The next theorem states that every subgroup of $\mathbf{Z}$ is of this form.

Theorem Suppose $G \subset \mathbf{Z}$ is a subgroup. Then

1) $0 \in G$.
2) If g_1 and $g_2 \in G$, then $(m_1g_1 + m_2g_2) \in G$ for all integers m_1, m_2.
3) $\exists!$ non-negative integer n such that $G = n\mathbf{Z}$. In fact, if $G \neq \{0\}$ and n is the smallest positive integer in G, then $G = n\mathbf{Z}$.

Proof Since G is non-void, $\exists\ g \in G$. Now $(-g) \in G$ and thus $0 = g + (-g)$ belongs to G, and so 1) is true. Part 2) is straightforward, so consider 3). If $G \neq 0$, it must contain a positive element. Let n be the smallest positive integer in G. If $g \in G$, $g = nm + r$ where $0 \leq r < n$. Since $r \in G$, it must be 0, and $g \in n\mathbf{Z}$.

Now suppose $a, b \in \mathbf{Z}$ and at least one of a and b is non-zero.

Theorem Let G be the set of all linear combinations of a and b, i.e., $G = \{ma + nb : m, n \in \mathbf{Z}\}$. Then

1) G contains a and b.
2) G is a subgroup. In fact, it is the smallest subgroup containing a and b. It is called the subgroup generated by a and b.
3) Denote by (a, b) the smallest positive integer in G. By the previous theorem, $G = (a, b)\mathbf{Z}$, and thus $(a, b) \mid a$ and $(a, b) \mid b$. Also note that $\exists\ m, n$ such that $ma + nb = (a, b)$. The integer (a, b) is called the *greatest common divisor* of a and b.
4) If n is an integer which divides a and b, then n also divides (a, b).

Proof of 4) Suppose $n \mid a$ and $n \mid b$ i.e., suppose $a, b \in n\mathbf{Z}$. Since G is the smallest subgroup containing a and b, $n\mathbf{Z} \supset (a, b)\mathbf{Z}$, and thus $n \mid (a, b)$.

Corollary The following are equivalent.

1) a and b have no common divisors, i.e., $(n \mid a \text{ and } n \mid b) \Rightarrow n = \pm 1$.

2) $(a, b) = 1$, i.e., the subgroup generated by a and b is all of **Z**.
3) $\exists\ m, n \in \mathbf{Z}$ with $ma + nb = 1$.

Definition If any one of these three conditions is satisfied, we say that a and b are *relatively prime.*

This next theorem is the basis for unique factorization.

Theorem If a and b are relatively prime with a not zero, then $a|bc \Rightarrow a|c$.

Proof Suppose a and b are relatively prime, $c \in \mathbf{Z}$ and $a \mid bc$. Then there exist m, n with $ma + nb = 1$, and thus $mac + nbc = c$. Now $a \mid mac$ and $a \mid nbc$. Thus $a \mid (mac + nbc)$ and so $a \mid c$.

Definition A *prime* is an integer $p > 1$ which does not factor, i.e., if $p = ab$ then $a = \pm 1$ or $a = \pm p$. The first few primes are 2, 3, 5, 7, 11, 13, 17,... .

Theorem Suppose p is a prime.

1) If a is an integer which is not a multiple of p, then $(p, a) = 1$. In other words, if a is any integer, $(p, a) = p$ or $(p, a) = 1$.
2) If $p \mid ab$ then $p \mid a$ or $p \mid b$.
3) If $p \mid a_1 a_2 \cdots a_n$ then p divides some a_i. Thus if each a_i is a prime, then p is equal to some a_i.

Proof Part 1) follows immediately from the definition of prime. Now suppose $p \mid ab$. If p does not divide a, then by 1), $(p, a) = 1$ and by the previous theorem, p must divide b. Thus 2) is true. Part 3) follows from 2) and induction on n.

The Unique Factorization Theorem Suppose a is an integer which is not 0,1, or -1. Then a may be factored into the product of primes and, except for order, this factorization is unique. That is, $\exists$ a unique collection of distinct primes $p_1, p_2, ..., p_k$ and positive integers $s_1, s_2, ..., s_k$ such that $a = \pm p_1^{s_1} p_2^{s_2} \cdots p_k^{s_k}$.

Proof Factorization into primes is obvious, and uniqueness follows from 3) in the theorem above. The power of this theorem is uniqueness, not existence.

Now that we have unique factorization and part 3) above, the picture becomes transparent. Here are some of the basic properties of the integers in this light.

Theorem (Summary)

1) Suppose $|a| > 1$ has prime factorization $a = \pm p_1^{s_1} \cdots p_k^{s_k}$. Then the only divisors of a are of the form $\pm p_1^{t_1} \cdots p_k^{t_k}$ where $0 \le t_i \le s_i$ for $i = 1, ..., k$.

2) If $| a | > 1$ and $| b | > 1$, then $(a, b) = 1$ iff there is no common prime in their factorizations. Thus if there is no common prime in their factorizations, $\exists\ m, n$ with $ma + nb = 1$, and also $(a^2, b^2) = 1$.

3) Suppose $|a| > 1$ and $|b| > 1$. Let $\{p_1, \ldots, p_k\}$ be the union of the distinct primes of their factorizations. Thus $a = \pm p_1^{s_1} \cdots p_k^{s_k}$ where $0 \le s_i$ and $b = \pm p_1^{t_1} \cdots p_k^{t_k}$ where $0 \le t_i$. Let u_i be the minimum of s_i and t_i. Then $(a, b) = p_1^{u_1} \cdots p_k^{u_k}$. For example $(2^3 \cdot 5 \cdot 11,\ 2^2 \cdot 5^4 \cdot 7) = 2^2 \cdot 5$.

3′) Let v_i be the maximum of s_i and t_i. Then $c = p_1^{v_1} \cdots p_k^{v_k}$ is the *least* (positive) *common multiple* of a and b. Note that c is a multiple of a and b, and if n is a multiple of a and b, then n is a multiple of c. Finally, if a and b are positive, their least common multiple is $c = ab/(a, b)$, and if in addition a and b are relatively prime, then their least common multiple is just their product.

4) There is an infinite number of primes. (Proof: Suppose there were only a finite number of primes $p_1, p_2, ..., p_k$. Then no prime would divide $(p_1 p_2 \cdots p_k + 1)$.)

5) Suppose c is an integer greater than 1. Then $\sqrt{c}$ is rational iff $\sqrt{c}$ is an integer. In particular, $\sqrt{2}$ and $\sqrt{3}$ are irrational. (Proof: If $\sqrt{c}$ is rational, $\exists$ positive integers a and b with $\sqrt{c} = a/b$ and $(a, b) = 1$. If $b > 1$, then it is divisible by some prime, and since $cb^2 = a^2$, this prime will also appear in the prime factorization of a. This is a contradiction and thus $b = 1$ and $\sqrt{c}$ is an integer.) (See the fifth exercise below.)

Exercise Find (180,28), i.e., find the greatest common divisor of 180 and 28, i.e., find the positive generator of the subgroup generated by {180,28}. Find integers m and n such that $180m + 28n = (180, 28)$. Find the least common multiple of 180 and 28, and show that it is equal to $(180 \cdot 28)/(180, 28)$.

Exercise We have defined the greatest common divisor (gcd) and the least common multiple (lcm) of a pair of integers. Now suppose $n \geq 2$ and $S = \{a_1, a_2, .., a_n\}$ is a finite collection of integers with $|a_i| > 1$ for $1 \leq i \leq n$. Define the gcd and the lcm of the elements of S and develop their properties. Express the gcd and the lcm in terms of the prime factorizations of the a_i. When is the lcm of S equal to the product $a_1 a_2 \cdots a_n$? Show that the set of all linear combinations of the elements of S is a subgroup of $\mathbf{Z}$, and its positive generator is the gcd of the elements of S.

Exercise Show that the gcd of $S = \{90, 70, 42\}$ is 2, and find integers n_1, n_2, n_3 such that $90n_1 + 70n_2 + 42n_3 = 2$. Also find the lcm of the elements of S.

Exercise Show that if each of $G_1, G_2, ..., G_m$ is a subgroup of $\mathbf{Z}$, then $G_1 \cap G_2 \cap \cdots \cap G_m$ is also a subgroup of $\mathbf{Z}$. Now let $G = (90\mathbf{Z}) \cap (70\mathbf{Z}) \cap (42\mathbf{Z})$ and find the positive integer n with $G = n\mathbf{Z}$.

Exercise Show that if the nth root of an integer is a rational number, then it itself is an integer. That is, suppose c and n are integers greater than 1. There is a unique positive real number x with $x^n = c$. Show that if x is rational, then it is an integer. Thus if p is a prime, its nth root is an irrational number.

Exercise Show that a positive integer is divisible by 3 iff the sum of its digits is divisible by 3. More generally, let $a = a_n a_{n-1} \ldots a_0 = a_n 10^n + a_{n-1} 10^{n-1} + \cdots + a_0$ where $0 \leq a_i \leq 9$. Now let $b = a_n + a_{n-1} + \cdots + a_0$, and show that 3 divides a and b with the same remainder. Although this is a straightforward exercise in long division, it will be more transparent later on. In the language of the next chapter, it says that $[a] = [b]$ in $\mathbf{Z}_3$.

Card Trick Ask friends to pick out seven cards from a deck and then to select one to look at without showing it to you. Take the six cards face down in your left hand and the selected card in your right hand, and announce you will place the selected card in with the other six, but they are not to know where. Put your hands behind your back and place the selected card on top, and bring the seven cards in front in your left hand. Ask your friends to give you a number between one and seven (not allowing one). Suppose they say three. You move the top card to the bottom, then the second card to the bottom, and then you turn over the third card, leaving it face up on top. Then repeat the process, moving the top two cards to the bottom and turning the third card face up on top. Continue until there is only one card face down, and this will be the selected card. Magic? Stay tuned for Chapter 2, where it is shown that any non-zero element of $\mathbf{Z}_7$ has order 7.

Chapter 2

Groups

Groups are the central objects of algebra. In later chapters we will define rings and modules and see that they are special cases of groups. Also ring homomorphisms and module homomorphisms are special cases of group homomorphisms. Even though the definition of group is simple, it leads to a rich and amazing theory. Everything presented here is standard, except that the product of groups is given in the additive notation. This is the notation used in later chapters for the products of rings and modules. This chapter and the next two chapters are restricted to the most basic topics. The approach is to do quickly the fundamentals of groups, rings, and matrices, and to push forward to the chapter on linear algebra. This chapter is, by far and above, the most difficult chapter in the book, because group operations may be written as addition or multiplication, and also the concept of coset is confusing at first.

Definition Suppose G is a non-void set and $\phi : G \times G \to G$ is a function. ϕ is called a *binary operation*, and we will write $\phi(a, b) = a \cdot b$ or $\phi(a, b) = a + b$. Consider the following properties.

1) If $a, b, c \in G$ then $a \cdot (b \cdot c) = (a \cdot b) \cdot c$.	If $a, b, c \in G$ then $a + (b + c) = (a + b) + c$.
2) $\exists\ e = e_G \in G$ such that if $a \in G$ $e \cdot a = a \cdot e = a$.	$\exists\ \underline{0} = \underline{0}_G \in G$ such that if $a \in G$ $\underline{0} + a = a + \underline{0} = a$.
3) If $a \in G$, $\exists b \in G$ with $a \cdot b = b \cdot a = e$ (b is written as $b = a^{-1}$).	If $a \in G, \exists b \in G$ with $a + b = b + a = \underline{0}$ (b is written as $b = -a$).
4) If $a, b \in G$, then $a \cdot b = b \cdot a$.	If $a, b \in G$, then $a + b = b + a$.

Definition If properties 1), 2), and 3) hold, (G, ϕ) is said to be a *group*. If we write $\phi(a, b) = a \cdot b$, we say it is a *multiplicative* group. If we write $\phi(a, b) = a + b$,

we say it is an *additive* group. If in addition, property 4) holds, we say the group is *abelian* or *commutative*.

Theorem Let (G, ϕ) be a multiplicative group.

(i) Suppose $a, c, \bar{c} \in G$. Then $a \cdot c = a \cdot \bar{c} \Rightarrow c = \bar{c}$.
Also $c \cdot a = \bar{c} \cdot a \Rightarrow c = \bar{c}$.
In other words, if $f : G \to G$ is defined by $f(c) = a \cdot c$, then f is injective. Also f is bijective with f^{-1} given by $f^{-1}(c) = a^{-1} \cdot c$.

(ii) e is unique, i.e., if $\bar{e} \in G$ satisfies 2), then $e = \bar{e}$. In fact, if $a, b \in G$ then $(a \cdot b = a) \Rightarrow (b = e)$ and $(a \cdot b = b) \Rightarrow (a = e)$. Recall that b is an identity in G provided it is a right and left identity for any a in G. However, group structure is so rigid that if $\exists\ a \in G$ such that b is a right identity for a, then $b = e$. Of course, this is just a special case of the cancellation law in (i).

(iii) Every right inverse is an inverse, i.e., if $a \cdot b = e$ then $b = a^{-1}$. Also if $b \cdot a = e$ then $b = a^{-1}$. Thus inverses are unique.

(iv) If $a \in G$, then $(a^{-1})^{-1} = a$.

(v) The multiplication $a_1 \cdot a_2 \cdot a_3 = a_1 \cdot (a_2 \cdot a_3) = (a_1 \cdot a_2) \cdot a_3$ is well-defined. In general, $a_1 \cdot a_2 \cdots a_n$ is well defined.

(vi) If $a, b \in G$, $(a \cdot b)^{-1} = b^{-1} \cdot a^{-1}$. Also $(a_1 \cdot a_2 \cdots a_n)^{-1} = a_n^{-1} \cdot a_{n-1}^{-1} \cdots a_1^{-1}$.

(vii) Suppose $a \in G$. Let $a^0 = e$ and if $n > 0$, $a^n = a \cdots a$ (n times) and $a^{-n} = a^{-1} \cdots a^{-1}$ (n times). If $n_1, n_2, ..., n_t \in \mathbf{Z}$ then $a^{n_1} \cdot a^{n_2} \cdots a^{n_t} = a^{n_1 + \cdots + n_t}$. Also $(a^n)^m = a^{nm}$. Finally, if G is abelian and $a, b \in G$, then $(a \cdot b)^n = a^n \cdot b^n$.

Exercise. Write out the above theorem where G is an additive group. Note that part (vii) states that G has a scalar multiplication over $\mathbf{Z}$. This means that if a is in G and n is an integer, there is defined an element an in G. This is so basic, that we state it explicitly.

Theorem. Suppose G is an additive group. If $a \in G$, let $a0 = \underline{0}$ and if $n > 0$, let $an = (a + \cdot \cdot + a)$ where the sum is n times, and $a(-n) = (-a) + (-a) \cdot \cdot + (-a)$,

which we write as $(-a - a \cdots - a)$. Then the following properties hold in general, except the first requires that G be abelian.

$$\begin{aligned} (a+b)n &= an+bn \\ a(n+m) &= an+am \\ a(nm) &= (an)m \\ a1 &= a \end{aligned}$$

Note that the plus sign is used ambiguously — sometimes for addition in G and sometimes for addition in $\mathbf{Z}$. In the language used in Chapter 5, this theorem states that any additive abelian group is a $\mathbf{Z}$-module. (See page 71.)

Exercise Suppose G is a non-void set with a binary operation $\phi(a,b) = a \cdot b$ which satisfies 1), 2) and [3′) If $a \in G$, $\exists b \in G$ with $a \cdot b = e$]. Show (G, ϕ) is a group, i.e., show $b \cdot a = e$. In other words, the group axioms are stronger than necessary. If every element has a right inverse, then every element has a two sided inverse.

Exercise Suppose G is the set of all functions from $\mathbf{Z}$ to $\mathbf{Z}$ with multiplication defined by composition, i.e., $f \cdot g = f \circ g$. Note that G satisfies 1) and 2) but not 3), and thus G is not a group. Show that f has a right inverse in G iff f is surjective, and f has a left inverse in G iff f is injective (see page 10). Also show that the set of all bijections from $\mathbf{Z}$ to $\mathbf{Z}$ is a group under composition.

Examples $G = \mathbf{R}$, $G = \mathbf{Q}$, or $G = \mathbf{Z}$ with $\phi(a,b) = a + b$ is an additive abelian group.

Examples $G = \mathbf{R} - 0$ or $G = \mathbf{Q} - 0$ with $\phi(a,b) = ab$ is a multiplicative abelian group.
$G = \mathbf{Z} - 0$ with $\phi(a,b) = ab$ is not a group.
$G = \mathbf{R}^+ = \{r \in \mathbf{R} : r > 0\}$ with $\phi(a,b) = ab$ is a multiplicative abelian group.

Subgroups

Theorem Suppose G is a multiplicative group and $H \subset G$ is a non-void subset satisfying

1) if $a, b \in H$ then $a \cdot b \in H$

and 2) if $a \in H$ then $a^{-1} \in H$.

Then $e \in H$ and H is a group under multiplication. H is called a *subgroup* of G.

Proof Since H is non-void, $\exists a \in H$. By 2), $a^{-1} \in H$ and so by 1), $e \in H$. The associative law is immediate and so H is a group.

Example G is a subgroup of G and e is a subgroup of G. These are called the *improper* subgroups of G.

Example If $G = \mathbf{Z}$ under addition, and $n \in \mathbf{Z}$, then $H = n\mathbf{Z}$ is a subgroup of $\mathbf{Z}$. By a theorem in the section on the integers in Chapter 1, every subgroup of $\mathbf{Z}$ is of this form (see page 15). This is a key property of the integers.

Exercises Suppose G is a multiplicative group.

1) Let H be the *center* of G, i.e., $H = \{h \in G : g \cdot h = h \cdot g \text{ for all } g \in G\}$. Show H is a subgroup of G.

2) Suppose H_1 and H_2 are subgroups of G. Show $H_1 \cap H_2$ is a subgroup of G.

3) Suppose H_1 and H_2 are subgroups of G, with neither H_1 nor H_2 contained in the other. Show $H_1 \cup H_2$ is not a subgroup of G.

4) Suppose T is an index set and for each $t \in T$, H_t is a subgroup of G. Show $\bigcap_{t \in T} H_t$ is a subgroup of G.

5) Furthermore, if $\{H_t\}$ is a monotonic collection, then $\bigcup_{t \in T} H_t$ is a subgroup of G.

6) Suppose $G = \{\text{all functions } f : [0,1] \to \mathbf{R}\}$. Define an addition on G by $(f + g)(t) = f(t) + g(t)$ for all $t \in [0,1]$. This makes G into an abelian group. Let K be the subset of G composed of all differentiable functions. Let H be the subset of G composed of all continuous functions. What theorems in calculus show that H and K are subgroups of G? What theorem shows that K is a subset (and thus subgroup) of H?

Order Suppose G is a multiplicative group. If G has an infinite number of

elements, we say that $o(G)$, the *order* of G, is infinite. If G has n elements, then $o(G) = n$. Suppose $a \in G$ and $H = \{a^i : i \in \mathbf{Z}\}$. H is an abelian subgroup of G called the *subgroup generated by* a. We define the *order of the element* a to be the order of H, i.e., the order of the subgroup generated by a. Let $f : \mathbf{Z} \to H$ be the surjective function defined by $f(m) = a^m$. Note that $f(k + l) = f(k) \cdot f(l)$ where the addition is in $\mathbf{Z}$ and the multiplication is in the group H. We come now to the first real theorem in group theory. It says that the element a has finite order iff f is not injective, and in this case, the order of a is the smallest positive integer n with $a^n = e$.

Theorem Suppose a is an element of a multiplicative group G, and $H = \{a^i : i \in \mathbf{Z}\}$. If $\exists$ distinct integers i and j with $a^i = a^j$, then a has some finite order n. In this case H has n distinct elements, $H = \{a^0, a^1, \ldots, a^{n-1}\}$, and $a^m = e$ iff $n|m$. In particular, the order of a is the smallest positive integer n with $a^n = e$, and $f^{-1}(e) = n\mathbf{Z}$.

Proof Suppose $j < i$ and $a^i = a^j$. Then $a^{i-j} = e$ and thus $\exists$ a smallest positive integer n with $a^n = e$. This implies that the elements of $\{a^0, a^1, ..., a^{n-1}\}$ are distinct, and we must show they are all of H. If $m \in \mathbf{Z}$, the Euclidean algorithm states that $\exists$ integers q and r with $0 \le r < n$ and $m = nq + r$. Thus $a^m = a^{nq} \cdot a^r = a^r$, and so $H = \{a^0, a^1, ..., a^{n-1}\}$, and $a^m = e$ iff $n|m$. Later in this chapter we will see that f is a homomorphism from an additive group to a multiplicative group and that, in additive notation, H is isomorphic to $\mathbf{Z}$ or $\mathbf{Z}_n$.

Exercise Write out this theorem for G an additive group. To begin, suppose a is an element of an additive group G, and $H = \{ai : i \in \mathbf{Z}\}$.

Exercise Show that if G is a finite group of even order, then G has an odd number of elements of order 2. Note that e is the only element of order 1.

Definition A group G is *cyclic* if $\exists$ an element of G which generates G.

Theorem If G is cyclic and H is a subgroup of G, then H is cyclic.

Proof Suppose $G = \{a^i : i \in \mathbf{Z}\}$ is a cyclic group and H is a subgroup of G. If $H = e$, then H is cyclic, so suppose $H \neq e$. Now there is a smallest positive integer m with $a^m \in H$. If t is an integer with $a^t \in H$, then by the Euclidean algorithm, m divides t, and thus a^m generates H. Note that in the case G has finite order n, i.e., $G = \{a^0, a^1, \ldots, a^{n-1}\}$, then $a^n = e \in H$, and thus the positive integer m divides n. In either case, we have a clear picture of the subgroups of G. Also note that this theorem was proved on page 15 for the additive group $\mathbf{Z}$.

Cosets Suppose H is a subgroup of a group G. It will be shown below that H partitions G into right cosets. It also partitions G into left cosets, and in general these partitions are distinct.

Theorem If H is a subgroup of a multiplicative group G, then $a \sim b$ defined by $a \sim b$ iff $a \cdot b^{-1} \in H$ is an equivalence relation. If $a \in G$, $cl(a) = \{b \in G : a \sim b\} = \{h \cdot a : h \in H\} = Ha$. Note that $a \cdot b^{-1} \in H$ iff $b \cdot a^{-1} \in H$.

If H is a subgroup of an additive group G, then $a \sim b$ defined by $a \sim b$ iff $(a - b) \in H$ is an equivalence relation. If $a \in G$, $cl(a) = \{b \in G : a \sim b\} = \{h + a : h \in H\} = H + a$. Note that $(a - b) \in H$ iff $(b - a) \in H$.

Definition These equivalence classes are called *right cosets*. If the relation is defined by $a \sim b$ iff $b^{-1} \cdot a \in H$, then the equivalence classes are $cl(a) = aH$ and they are called *left cosets*. H is a left and right coset. If G is abelian, there is no distinction between right and left cosets. Note that $b^{-1} \cdot a \in H$ iff $a^{-1} \cdot b \in H$.

In the theorem above, H is used to define an equivalence relation on G, and thus a partition of G. We now do the same thing a different way. We define the right cosets directly and show they form a partition of G. You might find this easier.

Theorem Suppose H is a subgroup of a multiplicative group G. If $a \in G$, define the right coset containing a to be $Ha = \{h \cdot a : h \in H\}$. Then the following hold.

1) $Ha = H$ iff $a \in H$.
2) If $b \in Ha$, then $Hb = Ha$, i.e., if $h \in H$, then $H(h \cdot a) = (Hh)a = Ha$.
3) If $Hc \cap Ha \neq \emptyset$, then $Hc = Ha$.
4) The right cosets form a partition of G, i.e., each a in G belongs to one and only one right coset.
5) Elements a and b belong to the same right coset iff $a \cdot b^{-1} \in H$ iff $b \cdot a^{-1} \in H$.

Proof There is no better way to develop facility with cosets than to prove this theorem. Also write this theorem for G an additive group.

Theorem Suppose H is a subgroup of a multiplicative group G.

1) Any two right cosets have the same number of elements. That is, if $a, b \in G$, $f : Ha \rightarrow Hb$ defined by $f(h \cdot a) = h \cdot b$ is a bijection. Also any two left cosets have the same number of elements. Since H is a right and left coset, any two cosets have the same number of elements.

2) G has the same number of right cosets as left cosets. The function F defined by $F(Ha) = a^{-1}H$ is a bijection from the collection of right cosets to the left cosets. The number of right (or left) cosets is called the *index* of H in G.

3) If G is finite, $o(H)$ (index of H) $= o(G)$ and so $o(H) \mid o(G)$. In other words, $o(G)/o(H) =$ the number of right cosets $=$ the number of left cosets.

4) If G is finite, and $a \in G$, then $o(a) \mid o(G)$. (Proof: The order of a is the order of the subgroup generated by a, and by 3) this divides the order of G.)

5) If G has prime order, then G is cyclic, and any element (except e) is a generator. (Proof: Suppose $o(G) = p$ and $a \in G$, $a \neq e$. Then $o(a) \mid p$ and thus $o(a) = p$.)

6) If $o(G) = n$ and $a \in G$, then $a^n = e$. (Proof: $a^{o(a)} = e$ and $n = o(a)\left(o(G)/o(a)\right)$.)

Exercises

i) Suppose G is a cyclic group of order 4, $G = \{e, a, a^2, a^3\}$ with $a^4 = e$. Find the order of each element of G. Find all the subgroups of G.

ii) Suppose G is the additive group $\mathbf{Z}$ and $H = 3\mathbf{Z}$. Find the cosets of H.

iii) Think of a circle as the interval $[0, 1]$ with end points identified. Suppose $G = \mathbf{R}$ under addition and $H = \mathbf{Z}$. Show that the collection of all the cosets of H can be thought of as a circle.

iv) Let $G = \mathbf{R}^2$ under addition, and H be the subgroup defined by $H = \{(a, 2a) : a \in \mathbf{R}\}$. Find the cosets of H. (See the last exercise on p 5.)

Normal Subgroups

We would like to make a group out of the collection of cosets of a subgroup H. In

general, there is no natural way to do that. However, it is easy to do in case H is a normal subgroup, which is described below.

Theorem If H is a subgroup of a group G, then the following are equivalent.

1) If $a \in G$, then $aHa^{-1} = H$
2) If $a \in G$, then $aHa^{-1} \subset H$
3) If $a \in G$, then $aH = Ha$
4) Every right coset is a left coset, i.e., if $a \in G$, $\exists\, b \in G$ with $Ha = bH$.

Proof 1) $\Rightarrow$ 2) is obvious. Suppose 2) is true and show 3). We have $(aHa^{-1})a \subset Ha$ so $aH \subset Ha$. Also $a(a^{-1}Ha) \subset aH$ so $Ha \subset aH$. Thus $aH = Ha$.
3) $\Rightarrow$ 4) is obvious. Suppose 4) is true and show 3). $Ha = bH$ contains a, so $bH = aH$ because a coset is an equivalence class. Thus $aH = Ha$.
Finally, suppose 3) is true and show 1). Multiply $aH = Ha$ on the right by a^{-1}.

Definition If H satisfies any of the four conditions above, then H is said to be a *normal* subgroup of G. (This concept goes back to Evariste Galois in 1831.)

Note For any group G, G and e are normal subgroups. If G is an abelian group, then every subgroup of G is normal.

Exercise Show that if H is a subgroup of G with index 2, then H is normal.

Exercise Show the intersection of a collection of normal subgroups of G is a normal subgroup of G. Show the union of a monotonic collection of normal subgroups of G is a normal subgroup of G.

Exercise Let $A \subset \mathbf{R}^2$ be the square with vertices $(-1,1), (1,1), (1,-1)$, and $(-1,-1)$, and G be the collection of all "isometries" of A onto itself. These are bijections of A onto itself which preserve distance and angles, i.e., which preserve dot product. Show that with multiplication defined as composition, G is a multiplicative group. Show that G has four rotations, two reflections about the axes, and two reflections about the diagonals, for a total of eight elements. Show the collection of rotations is a cyclic subgroup of order four which is a normal subgroup of G. Show that the reflection about the x-axis together with the identity form a cyclic subgroup of order two which is not a normal subgroup of G. Find the four right cosets of this subgroup. Finally, find the four left cosets of this subgroup.

Quotient Groups Suppose N is a normal subgroup of G, and C and D are cosets. We wish to define a coset E which is the product of C and D. If $c \in C$ and $d \in D$, define E to be the coset containing $c \cdot d$, i.e., $E = N(c \cdot d)$. The coset E does not depend upon the choice of c and d. This is made precise in the next theorem, which is quite easy.

Theorem Suppose G is a multiplicative group, N is a normal subgroup, and G/N is the collection of all cosets. Then $(Na) \cdot (Nb) = N(a \cdot b)$ is a well defined multiplication (binary operation) on G/N, and with this multiplication, G/N is a group. Its identity is N and $(Na)^{-1} = (Na^{-1})$. Furthermore, if G is finite, $o(G/N) = o(G)/o(N)$.

Proof Multiplication of elements in G/N is multiplication of subsets in G. $(Na) \cdot (Nb) = N(aN)b = N(Na)b = N(a \cdot b)$. Once multiplication is well defined, the group axioms are immediate.

Exercise Write out the above theorem for G an additive group. In the additive abelian group $\mathbf{R}/\mathbf{Z}$, determine those elements of finite order.

Example Suppose $G = \mathbf{Z}$ under $+$, $n > 1$, and $N = n\mathbf{Z}$. $\mathbf{Z}_n$, the *group of integers mod n* is defined by $\mathbf{Z}_n = \mathbf{Z}/n\mathbf{Z}$. If a is an integer, the coset $a + n\mathbf{Z}$ is denoted by $[a]$. Note that $[a] + [b] = [a + b]$, $-[a] = [-a]$, and $[a] = [a + nl]$ for any integer l. Any additive abelian group has a scalar multiplication over $\mathbf{Z}$, and in this case it is just $[a]m = [am]$. Note that $[a] = [r]$ where r is the remainder of a divided by n, and thus the distinct elements of $\mathbf{Z}_n$ are $[0], [1], ..., [n-1]$. Also $\mathbf{Z}_n$ is cyclic because each of $[1]$ and $[-1] = [n-1]$ is a generator. We already know that if p is a prime, any non-zero element of $\mathbf{Z}_p$ is a generator, because $\mathbf{Z}_p$ has p elements.

Theorem If $n > 1$ and a is any integer, then $[a]$ is a generator of $\mathbf{Z}_n$ iff $(a, n) = 1$.

Proof The element $[a]$ is a generator iff the subgroup generated by $[a]$ contains $[1]$ iff $\exists$ an integer k such that $[a]k = [1]$ iff $\exists$ integers k and l such that $ak + nl = 1$.

Exercise Show that a positive integer is divisible by 3 iff the sum of its digits is divisible by 3. Note that $[10] = [1]$ in $\mathbf{Z}_3$. (See the fifth exercise on page 18.)

Homomorphisms

Homomorphisms are functions between groups that commute with the group operations. It follows that they honor identities and inverses. In this section we list

the basic properties. Properties 11), 12), and 13) show the connections between coset groups and homomorphisms, and should be considered as the cornerstones of abstract algebra. As always, the student should rewrite the material in additive notation.

Definition If G and $\bar{G}$ are multiplicative groups, a function $f : G \to \bar{G}$ is a *homomorphism* if, for all $a, b \in G$, $f(a \cdot b) = f(a) \cdot f(b)$. On the left side, the group operation is in G, while on the right side it is in $\bar{G}$. The *kernel* of f is defined by $\ker(f) = f^{-1}(\bar{e}) = \{a \in G : f(a) = \bar{e}\}$. In other words, the kernel is the set of solutions to the equation $f(x) = \bar{e}$. (If $\bar{G}$ is an additive group, $\ker(f) = f^{-1}(\underline{0})$.)

Examples The constant map $f : G \to \bar{G}$ defined by $f(a) = \bar{e}$ is a homomorphism. If H is a subgroup of G, the inclusion $i : H \to G$ is a homomorphism. The function $f : \mathbf{Z} \to \mathbf{Z}$ defined by $f(t) = 2t$ is a homomorphism of additive groups, while the function defined by $f(t) = t + 2$ is not a homomorphism. The function $h : \mathbf{Z} \to \mathbf{R} - 0$ defined by $h(t) = 2^t$ is a homomorphism from an additive group to a multiplicative group.

We now catalog the basic properties of homomorphisms. These will be helpful later on in the study of ring homomorphisms and module homomorphisms.

Theorem Suppose G and $\bar{G}$ are groups and $f : G \to \bar{G}$ is a homomorphism.

1) $f(e) = \bar{e}$.

2) $f(a^{-1}) = f(a)^{-1}$. The first inverse is in G, and the second is in $\bar{G}$.

3) f is injective $\Leftrightarrow$ $\ker(f) = e$.

4) If H is a subgroup of G, $f(H)$ is a subgroup of $\bar{G}$. In particular, image(f) is a subgroup of $\bar{G}$.

5) If $\bar{H}$ is a subgroup of $\bar{G}$, $f^{-1}(\bar{H})$ is a subgroup of G. Furthermore, if $\bar{H}$ is normal in $\bar{G}$, then $f^{-1}(\bar{H})$ is normal in G.

6) The kernel of f is a normal subgroup of G.

7) If $\bar{g} \in \bar{G}$, $f^{-1}(\bar{g})$ is void or is a coset of $\ker(f)$, i.e., if $f(g) = \bar{g}$ then $f^{-1}(\bar{g}) = Ng$ where $N = \ker(f)$. In other words, if the equation $f(x) = \bar{g}$ has a

solution, then the set of all solutions is a coset of $N=\ker(f)$. This is a key fact which is used routinely in topics such as systems of equations and linear differential equations.

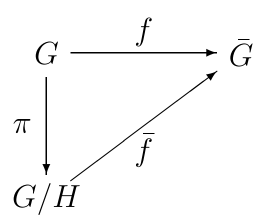
8) The composition of homomorphisms is a homomorphism, i.e., if $h : \bar{G} \to \bar{\bar{G}}$ is a homomorphism, then $h \circ f : G \to \bar{\bar{G}}$ is a homomorphism.

9) If $f : G \to \bar{G}$ is a bijection, then the function $f^{-1} : \bar{G} \to G$ is a homomorphism. In this case, f is called an *isomorphism*, and we write $G \approx \bar{G}$. In the case $G = \bar{G}$, f is also called an *automorphism*.

10) Isomorphisms preserve all algebraic properties. For example, if f is an isomorphism and $H \subset G$ is a subset, then H is a subgroup of G iff $f(H)$ is a subgroup of $\bar{G}$, H is normal in G iff $f(H)$ is normal in $\bar{G}$, G is cyclic iff $\bar{G}$ is cyclic, etc. Of course, this is somewhat of a cop-out, because an algebraic property is one that, by definition, is preserved under isomorphisms.

11) Suppose H is a normal subgroup of G. Then $\pi : G \to G/H$ defined by $\pi(a) = Ha$ is a surjective homomorphism with kernel H. Furthermore, if $f : G \to \bar{G}$ is a surjective homomorphism with kernel H, then $G/H \approx \bar{G}$ (see below).

12) Suppose H is a normal subgroup of G. If $H \subset \ker(f)$, then $\bar{f} : G/H \to \bar{G}$ defined by $\bar{f}(Ha) = f(a)$ is a well-defined homomorphism making the following diagram commute.



Thus defining a homomorphism on a quotient group is the same as defining a homomorphism on the numerator which sends the denominator to $\bar{e}$. The image of $\bar{f}$ is the image of f and the kernel of $\bar{f}$ is $\ker(f)/H$. Thus if $H = \ker(f)$, $\bar{f}$ is injective, and thus $G/H \approx \text{image}(f)$.

13) Given any group homomorphism f, $\text{domain}(f)/\ker(f) \approx \text{image}(f)$. This is the fundamental connection between quotient groups and homomorphisms.

14) Suppose K is a group. Then K is an infinite cycle group iff K is isomorphic to the integers under addition, i.e., $K \approx \mathbf{Z}$. K is a cyclic group of order n iff $K \approx \mathbf{Z}_n$.

Proof of 14) Suppose $\bar{G} = K$ is generated by some element a. Then $f : \mathbf{Z} \to K$ defined by $f(m) = a^m$ is a homomorphism from an additive group to a multiplicative group. If $o(a)$ is infinite, f is an isomorphism. If $o(a) = n$, $\ker(f) = n\mathbf{Z}$ and $\bar{f} : \mathbf{Z}_n \to K$ is an isomorphism.

Exercise If a is an element of a group G, there is always a homomorphism from $\mathbf{Z}$ to G which sends 1 to a. When is there a homomorphism from $\mathbf{Z}_n$ to G which sends [1] to a? What are the homomorphisms from $\mathbf{Z}_2$ to $\mathbf{Z}_6$? What are the homomorphisms from $\mathbf{Z}_4$ to $\mathbf{Z}_8$?

Exercise Suppose G is a group and g is an element of G, $g \neq e$.

1) Under what conditions on g is there a homomorphism $f : \mathbf{Z}_7 \to G$ with $f([1]) = g$?

2) Under what conditions on g is there a homomorphism $f : \mathbf{Z}_{15} \to G$ with $f([1]) = g$?

3) Under what conditions on G is there an injective homomorphism $f : \mathbf{Z}_{15} \to G$?

4) Under what conditions on G is there a surjective homomorphism $f : \mathbf{Z}_{15} \to G$?

Exercise We know every finite group of prime order is cyclic and thus abelian. Show that every group of order four is abelian.

Exercise Let $G = \{h : [0,1] \to \mathbf{R} : h \text{ has an infinite number of derivatives}\}$. Then G is a group under addition. Define $f : G \to G$ by $f(h) = \frac{dh}{dt} = h'$. Show f is a homomorphism and find its kernel and image. Let $g : [0,1] \to \mathbf{R}$ be defined by $g(t) = t^3 - 3t + 4$. Find $f^{-1}(g)$ and show it is a coset of $\ker(f)$.

Exercise Let G be as above and $g \in G$. Define $f : G \to G$ by $f(h) = h'' + 5h' + 6t^2h$. Then f is a group homomorphism and the differential equation $h''+5h'+6t^2h = g$ has a solution iff g lies in the image of f. Now suppose this equation has a solution and $S \subset G$ is the set of all solutions. For which subgroup H of G is S an H-coset?

Exercise Suppose G is a multiplicative group and $a \in G$. Define $f : G \to G$ to be conjugation by a, i.e., $f(g) = a^{-1} \cdot g \cdot a$. Show that f is a homomorphism. Also show f is an automorphism and find its inverse.

Permutations

Suppose X is a (non-void) set. A bijection $f : X \to X$ is called a *permutation* on X, and the collection of all these permutations is denoted by $S = S(X)$. In this setting, variables are written on the left, i.e., $f = (x)f$. Therefore the composition $f \circ g$ means "f followed by g". $S(X)$ forms a multiplicative group under composition.

Exercise Show that if there is a bijection between X and Y, there is an isomorphism between $S(X)$ and $S(Y)$. Thus if each of X and Y has n elements, $S(X) \approx S(Y)$, and these groups are called the *symmetric* groups on n elements. They are all denoted by the one symbol S_n.

Exercise Show that $o(S_n) = n!$. Let $X = \{1, 2, ..., n\}$, $S_n = S(X)$, and $H = \{f \in S_n : (n)f = n\}$. Show H is a subgroup of S_n which is isomorphic to S_{n-1}. Let g be any permutation on X with $(n)g = 1$. Find $g^{-1}Hg$.

The next theorem shows that the symmetric groups are incredibly rich and complex.

Theorem (Cayley's Theorem) Suppose G is a multiplicative group with n elements and S_n is the group of all permutations on the set G. Then G is isomorphic to a subgroup of S_n.

Proof Let $h : G \to S_n$ be the function which sends a to the bijection $h_a : G \to G$ defined by $(g)h_a = g \cdot a$. The proof follows from the following observations.

1) For each given a, h_a is a bijection from G to G.
2) h is a homomorphism, i.e., $h_{a \cdot b} = h_a \circ h_b$.
3) h is injective and thus G is isomorphic to image$(h) \subset S_n$.

The Symmetric Groups Now let $n \geq 2$ and let S_n be the group of all permutations on $\{1, 2, ..., n\}$. The following definition shows that each element of S_n may

be represented by a matrix.

Definition Suppose $1 < k \leq n$, $\{a_1, a_2, ..., a_k\}$ is a collection of distinct integers with $1 \leq a_i \leq n$, and $\{b_1, b_2, ..., b_k\}$ is the same collection in some different order. Then the matrix $\begin{pmatrix} a_1 & a_2 \ldots a_k \\ b_1 & b_2 \ldots b_k \end{pmatrix}$ represents $f \in S_n$ defined by $(a_i)f = b_i$ for $1 \leq i \leq k$, and $(a)f = a$ for all other a. The composition of two permutations is computed by applying the matrix on the left first and the matrix on the right second.

There is a special type of permutation called a *cycle*. For these we have a special notation.

Definition $\begin{pmatrix} a_1 & a_2 ... a_{k-1} a_k \\ a_2 & a_3 ... a_k \quad a_1 \end{pmatrix}$ is called a k-cycle, and is denoted by $(a_1, a_2, ..., a_k)$. A 2-cycle is called a *transposition*. The cycles $(a_1, ..., a_k)$ and $(c_1, ..., c_\ell)$ are *disjoint* provided $a_i \neq c_j$ for all $1 \leq i \leq k$ and $1 \leq j \leq \ell$.

Listed here are eight basic properties of permutations. They are all easy except 4), which takes a little work. Properties 9) and 10) are listed solely for reference.

Theorem

1) Disjoint cycles commute. (This is obvious.)

2) Every nonidentity permutation can be written uniquely (except for order) as the product of disjoint cycles. (This is easy.)

3) Every permutation can be written (non-uniquely) as the product of transpositions. (Proof: $I = (1,2)(1,2)$ and $(a_1, ..., a_k) = (a_1, a_2)(a_1, a_3) \cdots (a_1, a_k)$.)

4) The parity of the number of these transpositions is unique. This means that if f is the product of p transpositions and also of q transpositions, then p is even iff q is even. In this case, f is said to be an *even* permutation. In the other case, f is an *odd* permutation.

5) A k-cycle is even (odd) iff k is odd (even). For example $(1,2,3) = (1,2)(1,3)$ is an even permutation.

6) Suppose $f, g \in S_n$. If one of f and g is even and the other is odd, then $g \circ f$ is

odd. If f and g are both even or both odd, then $g \circ f$ is even. (Obvious.)

7) The map $h : S_n \to \mathbf{Z}_2$ defined by $h(\text{even}) = [0]$ and $h(\text{odd}) = [1]$ is a homomorphism from a multiplicative group to an additive group. Its kernel (the subgroup of even permutations) is denoted by A_n and is called the *alternating* group. Thus A_n is a normal subgroup of index 2, and $S_n/A_n \approx \mathbf{Z}_2$.

8) If a, b, c and d are distinct integers in $\{1, 2, \ldots, n\}$, then $(a, b)(b, c) = (a, c, b)$ and $(a, b)(c, d) = (a, c, d)(a, c, b)$. Since $I = (1, 2, 3)^3$, it follows that for $n \geq 3$, every even permutation is the product of 3-cycles.

The following parts are not included in this course. They are presented here merely for reference.

9) For any $n \neq 4$, A_n is simple, i.e., has no proper normal subgroups.

10) S_n can be generated by two elements. In fact, $\{(1, 2), (1, 2, ..., n)\}$ generates S_n. (Of course there are subgroups of S_n which cannot be generated by two elements).

Proof of 4) It suffices to prove if the product of t transpositions is the identity I on $\{1, 2, \ldots, n\}$, then t is even. Suppose this is false and I is written as t transpositions, where t is the smallest odd integer this is possible. Since t is odd, it is at least 3. Suppose for convenience the first transposition is (a, n). We will rewrite I as a product of transpositions $\sigma_1 \sigma_2 \cdots \sigma_t$ where $(n)\sigma_i = (n)$ for $1 \leq i < t$ and $(n)\sigma_t \neq n$, which will be a contradiction. This can be done by inductively "pushing n to the right" using the equations below. If $a, b,$ and c are distinct integers in $\{1, 2, \ldots, n-1\}$, then $(a, n)(a, n) = I$, $(a, n)(b, n) = (a, b)(a, n)$, $(a, n)(a, c) = (a, c)(c, n)$, and $(a, n)(b, c) = (b, c)(a, n)$. Note that $(a, n)(a, n)$ cannot occur here because it would result in a shorter odd product. (Now you may solve the tile puzzle on page viii.)

Exercise

1) Write $\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 6 & 5 & 4 & 3 & 1 & 7 & 2 \end{pmatrix}$ as the product of disjoint cycles.
Write (1,5,6,7)(2,3,4)(3,7,1) as the product of disjoint cycles.
Write (3,7,1)(1,5,6,7)(2,3,4) as the product of disjoint cycles.
Which of these permutations are odd and which are even?

2) Suppose $(a_1, \ldots, a_k)$ and $(c_1, \ldots, c_\ell)$ are disjoint cycles. What is the order of their product?

3) Suppose $\sigma \in S_n$. Show that $\sigma^{-1}(1,2,3)\sigma = ((1)\sigma, (2)\sigma, (3)\sigma)$. This shows that conjugation by σ is just a type of relabeling. Also let $\tau = (4,5,6)$ and find $\tau^{-1}(1,2,3,4,5)\tau$.

4) Show that $H = \{\sigma \in S_6 : (6)\sigma = 6\}$ is a subgroup of S_6 and find its right cosets and its left cosets.

5) Let $A \subset \mathbf{R}^2$ be the square with vertices $(-1,1), (1,1), (1,-1)$, and $(-1,-1)$, and G be the collection of all isometries of A onto itself. We know from a previous exercise that G is a group with eight elements. It follows from Cayley's theorem that G is isomorphic to a subgroup of S_8. Show that G is isomorphic to a subgroup of S_4.

6) If G is a multiplicative group, define a new multiplication on the set G by $a \circ b = b \cdot a$. In other words, the new multiplication is the old multiplication in the opposite order. This defines a new group denoted by G^{op}, the opposite group. Show that it has the same identity and the same inverses as G, and that $f : G \to G^{op}$ defined by $f(a) = a^{-1}$ is a group isomorphism. Now consider the special case $G = S_n$. The convention used in this section is that an element of S_n is a permutation on $\{1, 2, \ldots, n\}$ with the variable written on the left. Show that an element of S_n^{op} is a permutation on $\{1, 2, \ldots, n\}$ with the variable written on the right. (Of course, either S_n or S_n^{op} may be called the symmetric group, depending on personal preference or context.)

Product of Groups

The product of groups is usually presented for multiplicative groups. It is presented here for additive groups because this is the form that occurs in later chapters. As an exercise, this section should be rewritten using multiplicative notation. The two theorems below are transparent and easy, but quite useful. For simplicity we first consider the product of two groups, although the case of infinite products is only slightly more difficult. For background, read first the two theorems on page 11.

Theorem Suppose G_1 and G_2 are additive groups. Define an addition on $G_1 \times G_2$ by $(a_1, a_2) + (b_1, b_2) = (a_1 + b_1, a_2 + b_2)$. This operation makes $G_1 \times G_2$ into a group. Its "zero" is $(\underline{0}_1, \underline{0}_2)$ and $-(a_1, a_2) = (-a_1, -a_2)$. The projections $\pi_1 : G_1 \times G_2 \to G_1$

and $\pi_2 : G_1 \times G_2 \to G_2$ are group homomorphisms. Suppose G is an additive group. We know there is a bijection from {functions $f : G \to G_1 \times G_2$} to {ordered pairs of functions (f_1, f_2) where $f_1 : G \to G_1$ and $f_2 : G \to G_2$}. Under this bijection, f is a group homomorphism iff each of f_1 and f_2 is a group homomorphism.

Proof It is transparent that the product of groups is a group, so let's prove the last part. Suppose G, G_1, and G_2 are groups and $f = (f_1, f_2)$ is a function from G to $G_1 \times G_2$. Now $f(a + b) = (f_1(a + b), f_2(a + b))$ and $f(a) + f(b) = (f_1(a), f_2(a)) + (f_1(b), f_2(b)) = (f_1(a) + f_1(b), f_2(a) + f_2(b))$. An examination of these two equations shows that f is a group homomorphism iff each of f_1 and f_2 is a group homomorphism.

Exercise Suppose G_1 and G_2 are groups. Show that $G_1 \times G_2$ and $G_2 \times G_1$ are isomorphic.

Exercise If $o(a_1) = m$ and $o(a_2) = n$, find the order of (a_1, a_2) in $G_1 \times G_2$.

Exercise Show that if G is any group of order 4, G is isomorphic to $\mathbf{Z}_4$ or $\mathbf{Z}_2 \times \mathbf{Z}_2$. Show $\mathbf{Z}_4$ is not isomorphic to $\mathbf{Z}_2 \times \mathbf{Z}_2$. Show $\mathbf{Z}_{12}$ is isomorphic to $\mathbf{Z}_4 \times \mathbf{Z}_3$. Finally, show that $\mathbf{Z}_{mn}$ is isomorphic to $\mathbf{Z}_m \times \mathbf{Z}_n$ iff $(m, n) = 1$.

Exercise Suppose G_1 and G_2 are groups and $i_1 : G_1 \to G_1 \times G_2$ is defined by $i_1(g_1) = (g_1, \underline{0}_2)$. Show i_1 is an injective group homomorphism and its image is a normal subgroup of $G_1 \times G_2$. Usually G_1 is identified with its image under i_1, so G_1 may be considered to be a normal subgroup of $G_1 \times G_2$. Let $\pi_2 : G_1 \times G_2 \to G_2$ be the projection map defined in the Background chapter. Show π_2 is a surjective homomorphism with kernel G_1. Therefore $(G_1 \times G_2)/G_1 \approx G_2$ as you would expect.

Exercise Let $\mathbf{R}$ be the reals under addition. Show that the addition in the product $\mathbf{R} \times \mathbf{R}$ is just the usual addition in analytic geometry.

Exercise Suppose $n > 2$. Is S_n isomorphic to $A_n \times G$ where G is a multiplicative group of order 2 ?

One nice thing about the product of groups is that it works fine for any finite number, or even any infinite number. The next theorem is stated in full generality.

Theorem Suppose T is an index set, and for any $t \in T$, G_t is an additive group. Define an addition on $\prod\limits_{t\in T} G_t = \prod G_t$ by $\{a_t\} + \{b_t\} = \{a_t + b_t\}$. This operation makes the product into a group. Its "zero" is $\{\underline{0}_t\}$ and $-\{a_t\} = \{-a_t\}$. Each projection $\pi_s : \prod G_t \to G_s$ is a group homomorphism. Suppose G is an additive group. Under the natural bijection from $\{$functions $f : G \to \prod G_t\}$ to $\{$sequences of functions $\{f_t\}_{t\in T}$ where $f_t : G \to G_t\}$, f is a group homomorphism iff each f_t is a group homomorphism. Finally, the scalar multiplication on $\prod G_t$ by integers is given coordinatewise, i.e., $\{a_t\}n = \{a_t n\}$.

Proof The addition on $\prod G_t$ is coordinatewise.

Exercise Suppose s is an element of T and $\pi_s : \prod G_t \to G_s$ is the projection map defined in the Background chapter. Show π_s is a surjective homomorphism and find its kernel.

Exercise Suppose s is an element of T and $i_s : G_s \to \prod G_t$ is defined by $i_s(a) = \{a_t\}$ where $a_t = \underline{0}$ if $t \neq s$ and $a_s = a$. Show i_s is an injective homomorphism and its image is a normal subgroup of $\prod G_t$. Thus each G_s may be considered to be a normal subgroup of $\prod G_t$.

Exercise Let $f : \mathbf{Z} \to \mathbf{Z}_{30} \times \mathbf{Z}_{100}$ be the homomorphism defined by $f(m) = ([4m], [3m])$. Find the kernel of f. Find the order of $([4], [3])$ in $\mathbf{Z}_{30} \times \mathbf{Z}_{100}$.

Exercise Let $f : \mathbf{Z} \to \mathbf{Z}_{90} \times \mathbf{Z}_{70} \times \mathbf{Z}_{42}$ be the group homomorphism defined by $f(m) = ([m], [m], [m])$. Find the kernel of f and show that f is not surjective. Let $g : \mathbf{Z} \to \mathbf{Z}_{45} \times \mathbf{Z}_{35} \times \mathbf{Z}_{21}$ be defined by $g(m) = ([m], [m], [m])$. Find the kernel of g and determine if g is surjective. Note that the gcd of $\{45, 35, 21\}$ is 1. Now let $h : \mathbf{Z} \to \mathbf{Z}_8 \times \mathbf{Z}_9 \times \mathbf{Z}_{35}$ be defined by $h(m) = ([m], [m], [m])$. Find the kernel of h and show that h is surjective. Finally suppose each of $b, c,$ and d is greater than 1 and $f : \mathbf{Z} \to \mathbf{Z}_b \times \mathbf{Z}_c \times \mathbf{Z}_d$ is defined by $f(m) = ([m], [m], [m])$. Find necessary and sufficient conditions for f to be surjective (see the first exercise on page 18).

Exercise Suppose T is a non-void set, G is an additive group, and G^T is the collection of all functions $f : T \to G$ with addition defined by $(f+g)(t) = f(t)+g(t)$. Show G^T is a group. For each $t \in T$, let $G_t = G$. Note that G^T is just another way of writing $\prod\limits_{t\in T} G_t$. Also note that if $T = [0, 1]$ and $G = \mathbf{R}$, the addition defined on G^T is just the usual addition of functions used in calculus. (For the ring and module versions, see exercises on pages 44 and 69.)

Chapter 3

Rings

Rings are additive abelian groups with a second operation called multiplication. The connection between the two operations is provided by the distributive law. Assuming the results of Chapter 2, this chapter flows smoothly. This is because ideals are also normal subgroups and ring homomorphisms are also group homomorphisms. We do not show that the polynomial ring $F[x]$ is a unique factorization domain, although with the material at hand, it would be easy to do. Also there is no mention of prime or maximal ideals, because these concepts are unnecessary for our development of linear algebra. These concepts are developed in the Appendix. A section on Boolean rings is included because of their importance in logic and computer science.

Suppose R is an additive abelian group, $R \neq \underline{0}$, and R has a second binary operation (i.e., map from $R \times R$ to R) which is denoted by multiplication. Consider the following properties.

1) If $a, b, c \in R$, $(a \cdot b) \cdot c = a \cdot (b \cdot c)$. (The associative property of multiplication.)
2) If $a, b, c \in R$, $a \cdot (b + c) = (a \cdot b) + (a \cdot c)$ and $(b + c) \cdot a = (b \cdot a) + (c \cdot a)$. (The distributive law, which connects addition and multiplication.)
3) R has a multiplicative identity, i.e., there is an element $\underline{1} = \underline{1}_R \in R$ such that if $a \in R$, $a \cdot \underline{1} = \underline{1} \cdot a = a$.
4) If $a, b \in R$, $a \cdot b = b \cdot a$. (The commutative property for multiplication.)

Definition If 1), 2), and 3) are satisfied, R is said to be a *ring*. If in addition 4) is satisfied, R is said to be a *commutative ring*.

Examples The basic commutative rings in mathematics are the integers **Z**, the

rational numbers $\mathbf{Q}$, the real numbers $\mathbf{R}$, and the complex numbers $\mathbf{C}$. It will be shown later that $\mathbf{Z}_n$, the integers mod n, has a natural multiplication under which it is a commutative ring. Also if R is any commutative ring, we will define $R[x_1, x_2, \ldots, x_n]$, a polynomical ring in n variables. Now suppose R is any ring, $n \geq 1$, and R_n is the collection of all $n \times n$ matrices over R. In the next chapter, operations of addition and multiplication of matrices will be defined. Under these operations, R_n is a ring. This is a basic example of a non-commutative ring. If $n > 1$, R_n is never commutative, even if R is commutative.

The next two theorems show that ring multiplication behaves as you would wish it to. They should be worked as exercises.

Theorem Suppose R is a ring and $a,\ b \in R$.

1) $a \cdot \underline{0} = \underline{0} \cdot a = \underline{0}$. Since $R \neq \underline{0}$, it follows that $\underline{1} \neq \underline{0}$.
2) $(-a) \cdot b = a \cdot (-b) = -(a \cdot b)$.

Recall that, since R is an additive abelian group, it has a scalar multiplication over $\mathbf{Z}$ (page 20). This scalar multiplication can be written on the right or left, i.e., $na = an$, and the next theorem shows it relates nicely to the ring multiplication.

Theorem Suppose $a, b \in R$ and $n, m \in \mathbf{Z}$.

1) $(na) \cdot (mb) = (nm)(a \cdot b)$. (This follows from the distributive law and the previous theorem.)

2) Let $\underline{n} = n\underline{1}$. For example, $\underline{2} = \underline{1} + \underline{1}$. Then $na = \underline{n} \cdot a$, that is, scalar multiplication by n is the same as ring multiplication by $\underline{n}$. Of course, $\underline{n}$ may be $\underline{0}$ even though $n \neq 0$.

Units

Definition An element a of a ring R is a *unit* provided $\exists$ an element $a^{-1} \in R$ with $a \cdot a^{-1} = a^{-1} \cdot a = \underline{1}$.

Theorem $\underline{0}$ can never be a unit. $\underline{1}$ is always a unit. If a is a unit, a^{-1} is also a unit with $(a^{-1})^{-1} = a$. The product of units is a unit with $(a \cdot b)^{-1} = b^{-1} \cdot a^{-1}$. More

generally, if $a_1, a_2, ..., a_n$ are units, then their product is a unit with $(a_1 \cdot a_2 \cdots a_n)^{-1} = a_n^{-1} \cdot a_{n-1}^{-1} \cdots a_1^{-1}$. The set of all units of R forms a multiplicative group denoted by R^*. Finally if a is a unit, $(-a)$ is a unit and $(-a)^{-1} = -(a^{-1})$.

In order for a to be a unit, it must have a two-sided inverse. It suffices to require a left inverse and a right inverse, as shown in the next theorem.

Theorem Suppose $a \in R$ and $\exists$ elements b and c with $b \cdot a = a \cdot c = \underline{1}$. Then $b = c$ and so a is a unit with $a^{-1} = b = c$.

Proof $b = b \cdot \underline{1} = b \cdot (a \cdot c) = (b \cdot a) \cdot c = \underline{1} \cdot c = c$.

Corollary Inverses are unique.

Domains and Fields In order to define these two types of rings, we first consider the concept of zero divisor.

Definition Suppose R is a commutative ring. An element $a \in R$ is called a *zero divisor* provided it is non-zero and $\exists$ a non-zero element b with $a \cdot b = \underline{0}$. Note that if a is a unit, it cannot be a zero divisor.

Theorem Suppose R is a commutative ring and $a \in (R - \underline{0})$ is not a zero divisor. Then $(a \cdot b = a \cdot c) \Rightarrow b = c$. In other words, multiplication by a is an injective map from R to R. It is surjective iff a is a unit.

Definition A *domain* (or *integral domain*) is a commutative ring such that, if $a \neq \underline{0}$, a is not a zero divisor. A *field* is a commutative ring such that, if $a \neq \underline{0}$, a is a unit. In other words, R is a field if it is commutative and its non-zero elements form a group under multiplication.

Theorem A field is a domain. A finite domain is a field.

Proof A field is a domain because a unit cannot be a zero divisor. Suppose R is a finite domain and $a \neq \underline{0}$. Then $f : R \to R$ defined by $f(b) = a \cdot b$ is injective and, by the pigeonhole principle, f is surjective. Thus a is a unit and so R is a field.

Exercise Let $\mathbf{C}$ be the additive abelian group $\mathbf{R}^2$. Define multiplication by $(a, b) \cdot (c, d) = (ac - bd,\ ad + bc)$. Show $\mathbf{C}$ is a commutative ring which is a field. Note that $\underline{1} = (1, 0)$ and if $i = (0, 1)$, then $i^2 = -\underline{1}$.

Examples $\mathbf{Z}$ is a domain. $\mathbf{Q}$, $\mathbf{R}$, and $\mathbf{C}$ are fields.

The Integers Mod n

The concept of integers mod n is fundamental in mathematics. It leads to a neat little theory, as seen by the theorems below. However, the basic theory cannot be completed until the product of rings is defined. (See the Chinese Remainder Theorem on page 50.) We know from page 27 that $\mathbf{Z}_n$ is an additive abelian group.

Theorem Suppose $n > 1$. Define a multiplication on $\mathbf{Z}_n$ by $[a] \cdot [b] = [ab]$. This is a well defined binary operation which makes $\mathbf{Z}_n$ into a commutative ring.

Proof Since $[a + kn] \cdot [b + ln] = [ab + n(al + bk + kln)] = [ab]$, the multiplication is well-defined. The ring axioms are easily verified.

Theorem Suppose $n > 1$ and $a \in \mathbf{Z}$. Then the following are equivalent.

1) $[a]$ is a generator of the additive group $\mathbf{Z}_n$.
2) $(a, n) = 1$.
3) $[a]$ is a unit of the ring $\mathbf{Z}_n$.

Proof We already know from page 27 that 1) and 2) are equivalent. Recall that if b is an integer, $[a]b = [a] \cdot [b] = [ab]$. Thus 1) and 3) are equivalent, because each says $\exists$ an integer b with $[a]b = [1]$.

Corollary If $n > 1$, the following are equivalent.

1) $\mathbf{Z}_n$ is a domain.
2) $\mathbf{Z}_n$ is a field.
3) n is a prime.

Proof We already know 1) and 2) are equivalent, because $\mathbf{Z}_n$ is finite. Suppose 3) is true. Then by the previous theorem, each of $[1]$, $[2]$,...,$[n - 1]$ is a unit, and thus 2) is true. Now suppose 3) is false. Then $n = ab$ where $1 < a < n$, $1 < b < n$,

$[a][b] = [0]$, and thus $[a]$ is a zero divisor and 1) is false.

Exercise List the units and their inverses for $\mathbf{Z}_7$ and $\mathbf{Z}_{12}$. Show that $(\mathbf{Z}_7)^*$ is a cyclic group but $(\mathbf{Z}_{12})^*$ is not. Show that in $\mathbf{Z}_{12}$ the equation $x^2 = \underline{1}$ has four solutions. Finally show that if R is a domain, $x^2 = \underline{1}$ can have at most two solutions in R (see the first theorem on page 46).

Subrings Suppose S is a subset of a ring R. The statement that S is a *subring* of R means that S is a subgroup of the group R, $\underline{1} \in S$, and $(a, b \in S \Rightarrow a \cdot b \in S)$. Then clearly S is a ring and has the same multiplicative identity as R. Note that $\mathbf{Z}$ is a subring of $\mathbf{Q}$, $\mathbf{Q}$ is a subring of $\mathbf{R}$, and $\mathbf{R}$ is a subring of $\mathbf{C}$. Subrings do not play a role analogous to subgroups. That role is played by ideals, and an ideal is never a subring (unless it is the entire ring). Note that if S is a subring of R and $s \in S$, then s may be a unit in R but not in S. Note also that $\mathbf{Z}$ and $\mathbf{Z}_n$ have no proper subrings, and thus occupy a special place in ring theory, as well as in group theory.

Ideals and Quotient Rings

Ideals in ring theory play a role analagous to normal subgroups in group theory.

Definition A subset I of a ring R is a $\left\{\begin{array}{l}\text{left}\\ \text{right}\\ 2\text{−sided}\end{array}\right\}$ ideal provided it is a subgroup of the additive group R and if $a \in R$ and $b \in I$, then $\left\{\begin{array}{l} a \cdot b \in I \\ b \cdot a \in I \\ a \cdot b \text{ and } b \cdot a \in I \end{array}\right\}$. The word "ideal " means "2-sided ideal". Of course, if R is commutative, every right or left ideal is an ideal.

Theorem Suppose R is a ring.

1) R and $\underline{0}$ are ideals of R. These are called the *improper* ideals.

2) If $\{I_t\}_{t \in T}$ is a collection of right (left, 2-sided) ideals of R, then $\bigcap\limits_{t \in T} I_t$ is a right (left, 2-sided) ideal of R. (See page 22.)

3) Furthermore, if the collection is monotonic, then $\bigcup_{t \in T} I_t$ is a right (left, 2-sided) ideal of R.

4) If $a \in R$, $I = aR$ is a right ideal. Thus if R is commutative, aR is an ideal, called a *principal ideal*. Thus every subgroup of $\mathbf{Z}$ is a principal ideal, because it is of the form $n\mathbf{Z}$.

5) If R is a commutative ring and $I \subset R$ is an ideal, then the following are equivalent.
   i) $I = R$.
   ii) I contains some unit u.
   iii) I contains $\underline{1}$.

Exercise Suppose R is a commutative ring. Show that R is a field iff R contains no proper ideals.

The following theorem is just an observation, but it is in some sense the beginning of ring theory.

Theorem Suppose R is a ring and $I \subset R$ is an ideal, $I \neq R$. Since I is a normal subgroup of the additive group R, R/I is an additive abelian group. Multiplication of cosets defined by $(a + I) \cdot (b + I) = (ab + I)$ is well-defined and makes R/I a ring.

Proof $(a + I) \cdot (b + I) = a \cdot b + aI + Ib + II \ \subset \ a \cdot b + I$. Thus multiplication is well defined, and the ring axioms are easily verified. The multiplicative identity is $(\underline{1} + I)$.

Observation If $R = \mathbf{Z}$, $n > 1$, and $I = n\mathbf{Z}$, the ring structure on $\mathbf{Z}_n = \mathbf{Z}/n\mathbf{Z}$ is the same as the one previously defined.

Homomorphisms

Definition Suppose R and $\bar{R}$ are rings. A function $f : R \to \bar{R}$ is a *ring homomorphism* provided

1) f is a group homomorphism
2) $f(\underline{1}_R) = \underline{1}_{\bar{R}}$ and
3) if $a, b \in R$ then $f(a \cdot b) = f(a) \cdot f(b)$. (On the left, multiplication

is in R, while on the right multiplication is in $\bar{R}$.)

The *kernel* of f is the kernel of f considered as a group homomorphism, namely $\ker(f) = f^{-1}(\underline{0})$.

Here is a list of the basic properties of ring homomorphisms. Much of this work has already been done by the theorem in group theory on page 28.

Theorem Suppose each of R and $\bar{R}$ is a ring.

1) The identity map $I_R : R \to R$ is a ring homomorphism.

2) The zero map from R to $\bar{R}$ is not a ring homomorphism (because it does not send $\underline{1}_R$ to $\underline{1}_{\bar{R}}$).

3) The composition of ring homomorphisms is a ring homomorphism.

4) If $f : R \to \bar{R}$ is a bijection which is a ring homomorphism, then $f^{-1} : \bar{R} \to R$ is a ring homomorphism. Such an f is called a *ring isomorphism*. In the case $R = \bar{R}$, f is also called a *ring automorphism*.

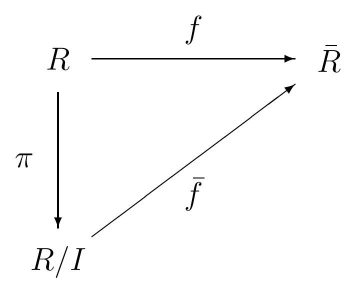
5) The image of a ring homomorphism is a subring of the range.

6) The kernel of a ring homomorphism is an ideal of the domain. In fact, if $f : R \to \bar{R}$ is a homomorphism and $I \subset \bar{R}$ is an ideal, then $f^{-1}(I)$ is an ideal of R.

7) Suppose I is an ideal of R, $I \neq R$, and $\pi : R \to R/I$ is the natural projection, $\pi(a) = (a + I)$. Then π is a surjective ring homomorphism with kernel I. Furthermore, if $f : R \to \bar{R}$ is a surjective ring homomorphism with kernel I, then $R/I \approx \bar{R}$ (see below).

8) From now on the word "homomorphism" means "ring homomorphism". Suppose $f : R \to \bar{R}$ is a homomorphism and I is an ideal of R, $I \neq R$. If $I \subset \ker(f)$, then $\bar{f} : R/I \to \bar{R}$ defined by $\bar{f}(a + I) = f(a)$

is a well-defined homomorphism making the following diagram commute.



Thus defining a homomorphism on a quotient ring is the same as defining a homomorphism on the numerator which sends the denominator to zero. The image of $\bar{f}$ is the image of f, and the kernel of $\bar{f}$ is $\ker(f)/I$. Thus if $I = \ker(f)$, $\bar{f}$ is injective, and so $R/I \approx$ image (f).

Proof We know all this on the group level, and it is only necessary to check that $\bar{f}$ is a ring homomorphism, which is obvious.

9) Given any ring homomorphism f, domain$(f)/\ker(f) \approx$ image(f).

Exercise Find a ring R with a proper ideal I and an element b such that b is not a unit in R but $(b + I)$ is a unit in R/I.

Exercise Show that if u is a unit in a ring R, then conjugation by u is an automorphism on R. That is, show that $f : R \to R$ defined by $f(a) = u^{-1} \cdot a \cdot u$ is a ring homomorphism which is an isomorphism.

Exercise Suppose T is a non-void set, R is a ring, and R^T is the collection of all functions $f : T \to R$. Define addition and multiplication on R^T point-wise. This means if f and g are functions from T to R, then $(f + g)(t) = f(t) + g(t)$ and $(f \cdot g)(t) = f(t)g(t)$. Show that under these operations R^T is a ring. Suppose S is a non-void set and $\alpha : S \to T$ is a function. If $f : T \to R$ is a function, define a function $\alpha^*(f) : S \to R$ by $\alpha^*(f) = f \circ \alpha$. Show $\alpha^* : R^T \to R^S$ is a ring homomorphism.

Exercise Now consider the case $T = [0, 1]$ and $R = \mathbf{R}$. Let $A \subset \mathbf{R}^{[0,1]}$ be the collection of all C^∞ functions, i.e., $A = \{f : [0, 1] \to \mathbf{R} : f$ has an infinite number of derivatives$\}$. Show A is a ring. Notice that much of the work has been done in the previous exercise. It is only necessary to show that A is a subring of the ring $\mathbf{R}^{[0,1]}$.

Polynomial Rings

In calculus, we consider real functions f which are polynomials, $f(x) = a_0 + a_1x + \cdot\cdot + a_nx^n$. The sum and product of polynomials are again polynomials, and it is easy to see that the collection of polynomial functions forms a commutative ring. We can do the same thing formally in a purely algebraic setting.

Definition Suppose R is a commutative ring and x is a "variable" or "symbol". The *polynomial ring* $R[x]$ is the collection of all polynomials $f = a_0 + a_1x + \cdot\cdot + a_nx^n$ where $a_i \in R$. Under the obvious addition and multiplication, $R[x]$ is a commutative ring. The *degree* of a non-zero polynomial f is the largest integer n such that $a_n \neq \underline{0}$, and is denoted by $n = \deg(f)$. If the top term $a_n = \underline{1}$, then f is said to be *monic*.

To be more formal, think of a polynomial $a_0 + a_1x + \cdots$ as an infinite sequence $(a_0, a_1, ...)$ such that each $a_i \in R$ and only a finite number are non-zero. Then
$(a_0, a_1, ...) + (b_0, b_1, ...) = (a_0 + b_0, a_1 + b_1, ...)$ and
$(a_0, a_1, ...) \cdot (b_0, b_1, ...) = (a_0b_0, a_0b_1 + a_1b_0, a_0b_2 + a_1b_1 + a_2b_0, ...)$.
Note that on the right, the ring multiplication $a \cdot b$ is written simply as ab, as is often done for convenience.

Theorem If R is a domain, $R[x]$ is also a domain.

Proof Suppose f and g are non-zero polynomials. Then $\deg(f)+\deg(g) = \deg(fg)$ and thus fg is not $\underline{0}$. Another way to prove this theorem is to look at the bottom terms instead of the top terms. Let a_ix^i and b_jx^j be the first non-zero terms of f and g. Then $a_ib_jx^{i+j}$ is the first non-zero term of fg.

Theorem (The Division Algorithm) Suppose R is a commutative ring, $f \in R[x]$ has degree ≥ 1 and its top coefficient is a unit in R. (If R is a field, the top coefficient of f will always be a unit.) Then for any $g \in R[x]$, $\exists!\ h, r \in R[x]$ such that $g = fh + r$ with $r = \underline{0}$ or $\deg(r) < \deg(f)$.

Proof This theorem states the existence and uniqueness of polynomials h and r. We outline the proof of existence and leave uniqueness as an exercise. Suppose $f = a_0 + a_1x + \cdot\cdot + a_mx^m$ where $m \geq 1$ and a_m is a unit in R. For any g with $\deg(g) < m$, set $h = \underline{0}$ and $r = g$. For the general case, the idea is to divide f into g until the remainder has degree less than m. The proof is by induction on the degree of g. Suppose $n \geq m$ and the result holds for any polynomial of degree less than

n. Suppose g is a polynomial of degree n. Now $\exists$ a monomial bx^t with $t = n - m$ and $\deg(g - fbx^t) < n$. By induction, $\exists$ h_1 and r with $fh_1 + r = (g - fbx^t)$ and $\deg(r) < m$. The result follows from the equation $f(h_1 + bx^t) + r = g$.

Note If $r = \underline{0}$ we say that f divides g. Note that $f = x - c$ divides g iff c is a *root* of g, i.e., $g(c) = \underline{0}$. More generally, $x - c$ divides g with remainder $g(c)$.

Theorem Suppose R is a domain, $n > 0$, and $g(x) = a_0 + a_1x + \cdots + a_nx^n$ is a polynomial of degree n with at least one root in R. Then g has at most n roots. Let $c_1, c_2, .., c_k$ be the distinct roots of g in the ring R. Then $\exists$ a unique sequence of positive integers $n_1, n_2, .., n_k$ and a unique polynomial h with no root in R so that $g(x) = (x - c_1)^{n_1} \cdots (x - c_k)^{n_k} h(x)$. (If h has degree 0, i.e., if $h = a_n$, then we say "all the roots of g belong to R". If $g = a_nx^n$, we say "all the roots of g are $\underline{0}$".)

Proof Uniqueness is easy so let's prove existence. The theorem is clearly true for $n = 1$. Suppose $n > 1$ and the theorem is true for any polynomial of degree less than n. Now suppose g is a polynomial of degree n and c_1 is a root of g. Then $\exists$ a polynomial h_1 with $g(x) = (x - c_1)h_1$. Since h_1 has degree less than n, the result follows by induction.

Note If g is any non-constant polynomial in $\mathbf{C}[x]$, all the roots of g belong to $\mathbf{C}$, i.e., $\mathbf{C}$ is an *algebraically closed field*. This is called The Fundamental Theorem of Algebra, and it is assumed without proof for this textbook.

Exercise Suppose g is a non-constant polynomial in $\mathbf{R}[x]$. Show that if g has odd degree then it has a real root. Also show that if $g(x) = x^2 + bx + c$, then it has a real root iff $b^2 \geq 4c$, and in that case both roots belong to $\mathbf{R}$.

Definition A domain T is a *principal ideal domain* (PID) if, given any ideal I, $\exists\ t \in T$ such that $I = tT$. Note that $\mathbf{Z}$ is a PID and any field is PID.

Theorem Suppose F is a field, I is a proper ideal of $F[x]$, and n is the smallest positive integer such that I contains a polynomial of degree n. Then I contains a unique polynomial of the form $f = a_0 + a_1x + \cdot\cdot + a_{n-1}x^{n-1} + x^n$ and it has the property that $I = fF[x]$. Thus $F[x]$ is a PID. Furthermore, each coset of I can be written uniquely in the form $(c_0 + c_1x + \cdot\cdot + c_{n-1}x^{n-1} + I)$.

Proof. This is a good exercise in the use of the division algorithm. Note this is similar to showing that a subgroup of $\mathbf{Z}$ is generated by one element (see page 15).

Theorem. Suppose R is a subring of a commutative ring C and $c \in C$. Then $\exists!$ homomorphism $h : R[x] \to C$ with $h(x) = c$ and $h(r) = r$ for all $r \in R$. It is defined by $h(a_0 + a_1x + \cdots + a_nx^n) = a_0 + a_1c + \cdots + a_nc^n$, i.e., h sends $f(x)$ to $f(c)$. The image of h is the smallest subring of C containing R and c.

This map h is called an *evaluation* map. The theorem says that adding two polynomials in $R[x]$ and evaluating is the same as evaluating and then adding in C. Also multiplying two polynomials in $R[x]$ and evaluating is the same as evaluating and then multiplying in C. In street language the theorem says you are free to send x wherever you wish and extend to a ring homomorphism on $R[x]$.

Exercise Let $\mathbf{C} = \{a + bi : a, b \in \mathbf{R}\}$. Since $\mathbf{R}$ is a subring of $\mathbf{C}$, there exists a homomorphism $h : \mathbf{R}[x] \to \mathbf{C}$ which sends x to i, and this h is surjective. Show $\ker(h) = (x^2 + 1)\mathbf{R}[x]$ and thus $\mathbf{R}[x]/(x^2 + 1) \approx \mathbf{C}$. This is a good way to look at the complex numbers, i.e., to obtain $\mathbf{C}$, adjoin x to $\mathbf{R}$ and set $x^2 = -1$.

Exercise $\mathbf{Z}_2[x]/(x^2 + x + 1)$ has 4 elements. Write out the multiplication table for this ring and show that it is a field.

Exercise Show that, if R is a domain, the units of $R[x]$ are just the units of R. Thus if F is a field, the units of $F[x]$ are the non-zero constants. Show that $[1] + [2]x$ is a unit in $\mathbf{Z}_4[x]$.

In this chapter we do not prove $F[x]$ is a unique factorization domain, nor do we even define unique factorization domain. The next definition and theorem are included merely for reference, and should not be studied at this stage.

Definition Suppose F is a field and $f \in F[x]$ has degree ≥ 1. The statement that g is an *associate* of f means $\exists$ a unit $u \in F[x]$ such that $g = uf$. The statement that f is *irreducible* means that if h is a non-constant polynomial which divides f, then h is an associate of f.

We do not develop the theory of $F[x]$ here. However, the development is easy because it corresponds to the development of $\mathbf{Z}$ in Chapter 1. The Division Algorithm corresponds to the Euclidean Algorithm. Irreducible polynomials correspond to prime integers. The degree function corresponds to the absolute value function. One difference is that the units of $F[x]$ are non-zero constants, while the units of $\mathbf{Z}$

are just ± 1. Thus the associates of f are all cf with $c \neq \underline{0}$ while the associates of an integer n are just $\pm n$. Here is the basic theorem. (This theory is developed in full in the Appendix under the topic of Euclidean domains.)

Theorem Suppose F is a field and $f \in F[x]$ has degree ≥ 1. Then f factors as the product of irreducibles, and this factorization is unique up to order and associates. Also the following are equivalent.

1) $F[x]/(f)$ is a domain.
2) $F[x]/(f)$ is a field.
3) f is irreducible.

Definition Now suppose x and y are "variables". If $a \in R$ and $n, m \geq 0$, then $ax^n y^m = ay^m x^n$ is called a *monomial*. Define an element of $R[x, y]$ to be any finite sum of monomials.

Theorem $R[x, y]$ is a commutative ring and $(R[x])[y] \approx R[x, y] \approx (R[y])[x]$. In other words, any polynomial in x and y with coefficients in R may be written as a polynomial in y with coefficients in $R[x]$, or as a polynomial in x with coefficients in $R[y]$.

Side Comment It is true that if F is a field, each $f \in F[x, y]$ factors as the product of irreducibles. However $F[x, y]$ is not a PID. For example, the ideal $I = xF[x, y] + yF[x, y] = \{f \in F[x, y] : f(\underline{0}, \underline{0}) = \underline{0}\}$ is not principal.

If R is a commutative ring and $n \geq 2$, the concept of a polynomial ring in n variables works fine without a hitch. If $a \in R$ and $v_1, v_2, ..., v_n$ are non-negative integers, then $ax_1^{v_1} x_2^{v_2} \cdots x_n^{v_n}$ is called a monomial. Order does not matter here. Define an element of $R[x_1, x_2, ..., x_n]$ to be any finite sum of monomials. This gives a commutative ring and there is canonical isomorphism $R[x_1, x_2, ..., x_n] \approx (R[x_1, x_2, ..., x_{n-1}])[x_n]$. Using this and induction on n, it is easy to prove the following theorem.

Theorem If R is a domain, $R[x_1, x_2, ..., x_n]$ is a domain and its units are just the units of R.

Exercise Suppose R is a commutative ring and $f : R[x,y] \to R[x]$ is the evaluation map which sends y to $\underline{0}$. This means $f(p(x,y)) = p(x,\underline{0})$. Show f is a ring homomorphism whose kernel is the ideal $(y) = yR[x,y]$. Use the fact that "the domain mod the kernel is isomorphic to the image" to show $R[x,y]/(y)$ is isomorphic to $R[x]$. That is, if you adjoin y to $R[x]$ and then factor it out, you get $R[x]$ back.

Product of Rings

The product of rings works fine, just as does the product of groups.

Theorem Suppose T is an index set and for each $t \in T$, R_t is a ring. On the additive abelian group $\prod_{t \in T} R_t = \prod R_t$, define multiplication by $\{r_t\} \cdot \{s_t\} = \{r_t \cdot s_t\}$. Then $\prod R_t$ is a ring and each projection $\pi_s : \prod R_t \to R_s$ is a ring homomorphism. Suppose R is a ring. Under the natural bijection from $\{$functions $f : R \to \prod R_t\}$ to $\{$sequences of functions $\{f_t\}_{t\in T}$ where $f_t : R \to R_t\}$, f is a ring homomorphism iff each f_t is a ring homomorphism.

Proof We already know f is a group homomorphism iff each f_t is a group homomorphism (see page 36). Note that $\{\underline{1}_t\}$ is the multiplicative identity of $\prod R_t$, and $f(\underline{1}_R) = \{\underline{1}_t\}$ iff $f_t(\underline{1}_R) = \underline{1}_t$ for each $t \in T$. Finally, since multiplication is defined coordinatewise, f is a ring homomorphism iff each f_t is a ring homomorphism.

Exercise Suppose R and S are rings. Note that $R \times 0$ is not a subring of $R \times S$ because it does not contain $(\underline{1}_R, \underline{1}_S)$. Show $R \times \underline{0}$ is an ideal and $(R \times S/R \times \underline{0}) \approx S$. Suppose $I \subset R$ and $J \subset S$ are ideals. Show $I \times J$ is an ideal of $R \times S$ and every ideal of $R \times S$ is of this form.

Exercise Suppose R and S are commutative rings. Show $T = R \times S$ is not a domain. Let $e = (\underline{1}, \underline{0}) \in R \times S$ and show $e^2 = e$, $(\underline{1} - e)^2 = (\underline{1} - e)$, $R \times \underline{0} = eT$, and $\underline{0} \times S = (\underline{1} - e)T$.

Exercise If T is any ring, an element e of T is called an *idempotent* provided $e^2 = e$. The elements $\underline{0}$ and $\underline{1}$ are idempotents called the *trivial* idempotents. Suppose T is a commutative ring and $e \in T$ is an idempotent with $\underline{0} \neq e \neq \underline{1}$. Let $R = eT$ and $S = (\underline{1} - e)T$. Show each of the ideals R and S is a ring with identity, and $f : T \to R \times S$ defined by $f(t) = (et, (\underline{1} - e)t)$ is a ring isomorphism. This shows that a commutative ring T splits as the product of two rings iff it contains a non-trivial idempotent.

The Chinese Remainder Theorem

The natural map from $\mathbf{Z}$ to $\mathbf{Z}_m \times \mathbf{Z}_n$ is a group homomorphism and also a ring homomorphism. If m and n are relatively prime, this map is surjective with kernel $mn\mathbf{Z}$, and thus $\mathbf{Z}_{mn}$ and $\mathbf{Z}_m \times \mathbf{Z}_n$ are isomorphic as groups and as rings. The next theorem is a classical generalization of this. (See exercise three on page 35.)

Theorem Suppose $n_1, ..., n_t$ are integers, each $n_i > 1$, and $(n_i, n_j) = 1$ for all $i \neq j$. Let $f_i : \mathbf{Z} \to \mathbf{Z}_{n_i}$ be defined by $f_i(a) = [a]$. (Note that the bracket symbol is used ambiguously.) Then the ring homomorphism $f = (f_1, .., f_t) : \mathbf{Z} \to \mathbf{Z}_{n_1} \times \cdot \cdot \times \mathbf{Z}_{n_t}$ is surjective. Furthermore, the kernel of f is $n\mathbf{Z}$, where $n = n_1 n_2 \cdot \cdot n_t$. Thus $\mathbf{Z}_n$ and $\mathbf{Z}_{n_1} \times \cdot \cdot \times \mathbf{Z}_{n_t}$ are isomorphic as rings, and thus also as groups.

Proof We wish to show that the order of $f(1)$ is n, and thus $f(1)$ is a group generator, and thus f is surjective. The element $f(1)m = ([1], .., [1])m = ([m], .., [m])$ is zero iff m is a multiple of each of $n_1, .., n_t$. Since their least common multiple is n, the order of $f(1)$ is n. (See the fourth exercise on page 36 for the case $t = 3$.)

Exercise Show that if a is an integer and p is a prime, then $[a] = [a^p]$ in $\mathbf{Z}_p$ (Fermat's Little Theorem). Use this and the Chinese Remainder Theorem to show that if b is a positive integer, it has the same last digit as b^5.

Characteristic

The following theorem is just an observation, but it shows that in ring theory, the ring of integers is a "cornerstone".

Theorem If R is a ring, there is one and only one ring homomorphism $f : \mathbf{Z} \to R$. It is given by $f(m) = m\underline{1} = \underline{m}$. Thus the subgroup of R generated by $\underline{1}$ is a subring of R isomorphic to $\mathbf{Z}$ or isomorphic to $\mathbf{Z}_n$ for some positive integer n.

Definition Suppose R is a ring and $f : \mathbf{Z} \to R$ is the natural ring homomorphism $f(m) = m\underline{1} = \underline{m}$. The non-negative integer n with $\ker(f) = n\mathbf{Z}$ is called the *characteristic* of R. Thus f is injective iff R has characteristic 0 iff $\underline{1}$ has infinite order. If f is not injective, the characteristic of R is the order of $\underline{1}$.

It is an interesting fact that, if R is a domain, all the non-zero elements of R have the same order. (See page 23 for the definition of order.)

Theorem Suppose R is a domain. If R has characteristic 0, then each non-zero $a \in R$ has infinite order. If R has finite characteristic n, then n is a prime and each non-zero $a \in R$ has order n.

Proof Suppose R has characteristic 0, a is a non-zero element of R, and m is a positive integer. Then $ma = \underline{m} \cdot a$ cannot be $\underline{0}$ because $\underline{m}, a \neq \underline{0}$ and R is a domain. Thus $o(a) = \infty$. Now suppose R has characteristic n. Then R contains $\mathbf{Z}_n$ as a subring, and thus $\mathbf{Z}_n$ is a domain and n is a prime. If a is a non-zero element of R, $na = \underline{n} \cdot a = \underline{0} \cdot a = \underline{0}$ and thus $o(a)|n$ and thus $o(a) = n$.

Exercise Show that if F is a field of characteristic 0, F contains $\mathbf{Q}$ as a subring. That is, show that the injective homomorphism $f : \mathbf{Z} \to F$ extends to an injective homomorphism $\bar{f} : \mathbf{Q} \to F$.

Boolean Rings

This section is not used elsewhere in this book. However it fits easily here, and is included for reference.

Definition A ring R is a *Boolean ring* if for each $a \in R$, $a^2 = a$, i.e., each element of R is an idempotent.

Theorem Suppose R is a Boolean ring.

1) R has characteristic 2. If $a \in R$, $2a = a + a = \underline{0}$, and so $a = -a$.

Proof $(a + a) = (a + a)^2 = a^2 + 2a^2 + a^2 = 4a$. Thus $2a = \underline{0}$.

2) R is commutative.

Proof $(a + b) = (a + b)^2 = a^2 + (a \cdot b) + (b \cdot a) + b^2$
$= a + (a \cdot b) - (b \cdot a) + b$. Thus $a \cdot b = b \cdot a$.

3) If R is a domain, $R \approx \mathbf{Z}_2$.

Proof Suppose $a \neq \underline{0}$. Then $a \cdot (\underline{1} - a) = \underline{0}$ and so $a = \underline{1}$.

4) The image of a Boolean ring is a Boolean ring. That is, if I is an ideal of R with $I \neq R$, then every element of R/I is idempotent and thus R/I is a Boolean ring. It follows from 3) that R/I is a domain iff R/I is a field iff $R/I \approx \mathbf{Z}_2$. (In the language of Chapter 6, I is a prime ideal iff I is a maximal ideal iff $R/I \approx \mathbf{Z}_2$).

Suppose X is a non-void set. If a is a subset of X, let $a' = (X - a)$ be a complement of a in X. Now suppose R is a non-void collection of subsets of X. Consider the following properties which the collection R may possess.

1) $a \in R \;\Rightarrow\; a' \in R$.
2) $a, b \in R \;\Rightarrow\; (a \cap b) \in R$.
3) $a, b \in R \;\Rightarrow\; (a \cup b) \in R$.
4) $\emptyset \in R$ and $X \in R$.

Theorem If 1) and 2) are satisfied, then 3) and 4) are satisfied. In this case, R is called a *Boolean algebra of sets*.

Proof Suppose 1) and 2) are true, and $a, b \in R$. Then $a \cup b = (a' \cap b')'$ belongs to R and so 3) is true. Since R is non-void, it contains some element a. Then $\emptyset = a \cap a'$ and $X = a \cup a'$ belong to R, and so 4) is true.

Theorem Suppose R is a Boolean algebra of sets. Define an addition on R by $a + b = (a \cup b) - (a \cap b)$. Under this addition, R is an abelian group with $\underline{0} = \emptyset$ and $a = -a$. Define a multiplication on R by $a \cdot b = a \cap b$. Under this multiplication R becomes a Boolean ring with $\underline{1} = X$.

Exercise Let $X = \{1, 2, ..., n\}$ and let R be the Boolean ring of all subsets of X. Note that $o(R) = 2^n$. Define $f_i : R \to \mathbf{Z}_2$ by $f_i(a) = [1]$ iff $i \in a$. Show each f_i is a homomorphism and thus $f = (f_1, ..., f_n) : R \to \mathbf{Z}_2 \times \mathbf{Z}_2 \times \cdot\cdot \times \mathbf{Z}_2$ is a ring homomorphism. Show f is an isomorphism. (See exercises 1) and 4) on page 12.)

Exercise Use the last exercise on page 49 to show that any finite Boolean ring is isomorphic to $\mathbf{Z}_2 \times \mathbf{Z}_2 \times \cdot\cdot \times \mathbf{Z}_2$, and thus also to the Boolean ring of subsets above.

Note Suppose R is a Boolean ring. It is a classical theorem that $\exists$ a Boolean algebra of sets whose Boolean ring is isomorphic to R. So let's just suppose R is a Boolean algebra of sets which is a Boolean ring with addition and multiplication defined as above. Now define $a \vee b = a \cup b$ and $a \wedge b = a \cap b$. These operations cup and cap are associative, commutative, have identity elements, and each distributes over the other. With these two operations (along with complement), R is called a *Boolean algebra*. R is not a group under cup or cap. Anyway, it is a classical fact that, if you have a Boolean ring (algebra), you have a Boolean algebra (ring). The advantage of the algebra is that it is symmetric in cup and cap. The advantage of the ring viewpoint is that you can draw from the rich theory of commutative rings.

Chapter 4

Matrices and Matrix Rings

We first consider matrices in full generality, i.e., over an arbitrary ring R. However, after the first few pages, it will be assumed that R is commutative. The topics, such as invertible matrices, transpose, elementary matrices, systems of equations, and determinant, are all classical. The highlight of the chapter is the theorem that a square matrix is a unit in the matrix ring iff its determinant is a unit in the ring. This chapter concludes with the theorem that similar matrices have the same determinant, trace, and characteristic polynomial. This will be used in the next chapter to show that an endomorphism on a finitely generated vector space has a well-defined determinant, trace, and characteristic polynomial.

Definition Suppose R is a ring and m and n are positive integers. Let $R_{m,n}$ be the collection of all $m \times n$ matrices

$$A = (a_{i,j}) = \begin{pmatrix} a_{1,1} & \dots & a_{1,n} \\ \vdots & & \vdots \\ a_{m,1} & \dots & a_{m,n} \end{pmatrix} \quad \text{where each entry} \;\; a_{i,j} \in R.$$

A matrix may be viewed as m n-dimensional row vectors or as n m-dimensional column vectors. A matrix is said to be *square* if it has the same number of rows as columns. Square matrices are so important that they have a special notation, $R_n = R_{n,n}$. R^n is defined to be the additive abelian group $R \times R \times \cdots \times R$. To emphasize that R^n does not have a ring structure, we use the "sum" notation, $R^n = R \oplus R \oplus \cdots \oplus R$. Our convention is to write elements of R^n as column vectors, i.e., to identify R^n with $R_{n,1}$. If the elements of R^n are written as row vectors, R^n is identified with $R_{1,n}$.

Addition of matrices To "add" two matrices, they must have the same number of rows and the same number of columns, i.e., addition is a binary operation $R_{m,n} \times R_{m,n} \to R_{m,n}$. The addition is defined by $(a_{i,j}) + (b_{i,j}) = (a_{i,j} + b_{i,j})$, i.e., the i, j term of the sum is the sum of the i, j terms. The following theorem is just an observation.

Theorem $R_{m,n}$ is an additive abelian group. Its "zero" is the matrix $\underline{0} = \underline{0}_{m,n}$ all of whose terms are zero. Also $-(a_{i,j}) = (-a_{i,j})$. Furthermore, as additive groups, $R_{m,n} \approx R^{mn}$.

Scalar multiplication An element of R is called a *scalar*. A matrix may be "multiplied" on the right or left by a scalar. Right scalar multiplication is defined by $(a_{i,j})c = (a_{i,j} \cdot c)$. It is a function $R_{m,n} \times R \to R_{m,n}$. Note in particular that scalar multiplication is defined on R^n. Of course, if R is commutative, there is no distinction between right and left scalar multiplication.

Theorem Suppose $A, B \in R_{m,n}$ and $c, d \in R$. Then

$$\begin{aligned} (A+B)c &= Ac + Bc \\ A(c+d) &= Ac + Ad \\ A(cd) &= (Ac)d \end{aligned}$$

and

$$A\underline{1} = A$$

This theorem is entirely transparent. In the language of the next chapter, it merely states that $R_{m,n}$ is a right module over the ring R.

Multiplication of Matrices The matrix product AB is defined iff the number of columns of A is equal to the number of rows of B. The matrix AB will have the same number of rows as A and the same number of columns as B, i.e., multiplication is a function $R_{m,n} \times R_{n,p} \to R_{m,p}$. The product $(a_{i,j})(b_{i,j})$ is defined to be the matrix whose (s,t) term is $a_{s,1} \cdot b_{1,t} + \cdots + a_{s,n} \cdot b_{n,t}$, i.e., the dot product of row s of A with column t of B.

Exercise Consider real matrices $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, $U = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$, $V = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, and $W = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$. Find the matrices AU, UA, AV, VA, AW, and WA.

Definition The *identity matrix* $I_n \in R_n$ is the square matrix whose diagonal terms are $\underline{1}$ and whose off-diagonal terms are $\underline{0}$.

Theorem Suppose $A \in R_{m,n}$.

1) $\underline{0}_{p,m}A = \underline{0}_{p,n} \qquad A\underline{0}_{n,p} = \underline{0}_{m,p}$
2) $I_m A = A = AI_n$

Theorem (The distributive laws) $(A + B)C = AC + BC$ and $C(A + B) = CA + CB$ whenever the operations are defined.

Theorem (The associative law for matrix multiplication) Suppose $A \in R_{m,n}$, $B \in R_{n,p}$, and $C \in R_{p,q}$. Then $(AB)C = A(BC)$. Note that $ABC \in R_{m,q}$.

Proof We must show that the (s,t) terms are equal. The proof involves writing it out and changing the order of summation. Let $(x_{i,j}) = AB$ and $(y_{i,j}) = BC$. Then the (s,t) term of $(AB)C$ is $\sum\limits_i x_{s,i}c_{i,t} = \sum\limits_i \Big(\sum\limits_j a_{s,j}b_{j,i}\Big) c_{i,t} = \sum\limits_{i,j} a_{s,j}b_{j,i}c_{i,t} = \sum\limits_j a_{s,j}\Big(\sum\limits_i b_{j,i}c_{i,t}\Big) = \sum\limits_j a_{s,j}y_{j,t}$ which is the (s,t) term of $A(BC)$.

Theorem For each ring R and integer $n \geq 1$, R_n is a ring.

Proof This elegant little theorem is immediate from the theorems above. The units of R_n are called *invertible* or *non-singular* matrices. They form a group under multiplication called the *general linear group* and denoted by $GL_n(R) = (R_n)^*$.

Exercise Recall that if A is a ring and $a \in A$, then aA is right ideal of A. Let $A = R_2$ and $a = (a_{i,j})$ where $a_{1,1} = \underline{1}$ and the other entries are $\underline{0}$. Find aR_2 and R_2a. Show that the only ideal of R_2 containing a is R_2 itself.

Multiplication by blocks Suppose $A, E \in R_n$, $B, F \in R_{n,m}$, $C, G \in R_{m,n}$, and $D, H \in R_m$. Then multiplication in R_{n+m} is given by

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}\begin{pmatrix} E & F \\ G & H \end{pmatrix} = \begin{pmatrix} AE + BG & AF + BH \\ CE + DG & CF + DH \end{pmatrix}.$$

Transpose

Notation *For the remainder of this chapter on matrices, suppose R is a commutative ring.* Of course, for $n > 1$, R_n is non-commutative.

Transpose is a function from $R_{m,n}$ to $R_{n,m}$. If $A \in R_{m,n}$, $A^t \in R_{n,m}$ is the matrix whose (i,j) term is the (j,i) term of A. So row i (column i) of A becomes column i (row i) of A^t. If A is an n-dimensional row vector, then A^t is an n-dimensional column vector. If A is a square matrix, A^t is also square.

Theorem 1) $(A^t)^t = A$

2) $(A+B)^t = A^t + B^t$

3) If $c \in R$, $(Ac)^t = A^t c$

4) $(AB)^t = B^t A^t$

5) If $A \in R_n$, then A is invertible iff A^t is invertible.
In this case $(A^{-1})^t = (A^t)^{-1}$.

Proof of 5) Suppose A is invertible. Then $I = I^t = (AA^{-1})^t = (A^{-1})^t A^t$.

Exercise Characterize those invertible matrices $A \in \mathbf{R}_2$ which have $A^{-1} = A^t$. Show that they form a subgroup of $GL_2(\mathbf{R})$.

Triangular Matrices

If $A \in R_n$, then A is *upper (lower) triangular* provided $a_{i,j} = \underline{0}$ for all $i > j$ (all $j > i$). A is *strictly upper (lower) triangular* provided $a_{i,j} = \underline{0}$ for all $i \geq j$ (all $j \geq i$). A is *diagonal* if it is upper and lower triangular, i.e., $a_{i,j} = \underline{0}$ for all $i \neq j$. Note that if A is upper (lower) triangular, then A^t is lower (upper) triangular.

Theorem If $A \in R_n$ is strictly upper (or lower) triangular, then $A^n = \underline{0}$.

Proof The way to understand this is just multiply it out for $n = 2$ and $n = 3$. The geometry of this theorem will become transparent later in Chapter 5 when the matrix A defines an R-module endomorphism on R^n (see page 93).

Definition If T is any ring, an element $t \in T$ is said to be *nilpotent* provided $\exists n$ such that $t^n = 0$. In this case, $(\underline{1} - t)$ is a unit with inverse $\underline{1} + t + t^2 + \cdots + t^{n-1}$. Thus if $T = R_n$ and B is a nilpotent matrix, $I - B$ is invertible.

Exercise Let $R = \mathbf{Z}$. Find the inverse of $\begin{pmatrix} 1 & 2 & -3 \\ 0 & 1 & 4 \\ 0 & 0 & 1 \end{pmatrix}$.

Exercise Suppose $A = \begin{pmatrix} a_1 & & & & \\ & a_2 & & 0 & \\ & & \cdot & & \\ & 0 & & \cdot & \\ & & & & a_n \end{pmatrix}$ is a diagonal matrix, $B \in R_{m,n}$, and $C \in R_{n,p}$. Show that BA is obtained from B by multiplying column i of B by a_i. Show AC is obtained from C by multiplying row i of C by a_i. Show A is a unit in R_n iff each a_i is a unit in R.

Scalar matrices A *scalar* matrix is a diagonal matrix for which all the diagonal terms are equal, i.e., a matrix of the form cI_n. The map $R \to R_n$ which sends c to cI_n is an injective ring homomorphism, and thus we may consider R to be a subring of R_n. Multiplying by a scalar is the same as multiplying by a scalar matrix, and thus scalar matrices commute with everything, i.e., if $B \in R_n$, $(cI_n)B = cB = Bc = B(cI_n)$. Recall we are assuming R is a commutative ring.

Exercise Suppose $A \in R_n$ and for each $B \in R_n$, $AB = BA$. Show A is a scalar matrix. For $n > 1$, this shows how non-commutative R_n is.

Elementary Operations and Elementary Matrices

Suppose R is a commutative ring and A is a matrix over R. There are 3 types of elementary row and column operations on the matrix A. A need not be square.

Type 1	Multiply row i by some unit $a \in R$.	Multiply column i by some unit $a \in R$.
Type 2	Interchange row i and row j.	Interchange column i and column j.
Type 3	Add a times row j to row i where $i \neq j$ and a is any element of R.	Add a times column i to column j where $i \neq j$ and a is any element of R.

Elementary Matrices Elementary matrices are square and invertible. There are three types. They are obtained by performing row or column operations on the identity matrix.

Type 1 $$B = \begin{pmatrix} 1 & & & & & \\ & 1 & & 0 & & \\ & & a & & & \\ & & & 1 & & \\ & 0 & & & 1 & \\ & & & & & 1 \end{pmatrix}$$ where a is a unit in R.

Type 2 $$B = \begin{pmatrix} 1 & & & & & \\ & 0 & & & 1 & \\ & & 1 & & & \\ & & & 1 & & \\ & 1 & & & 0 & \\ & & & & & 1 \end{pmatrix}$$

Type 3 $$B = \begin{pmatrix} 1 & & & & & \\ & 1 & & & a_{i,j} & \\ & & 1 & & & \\ & & & 1 & & \\ & 0 & & & 1 & \\ & & & & & 1 \end{pmatrix}$$ where $i \neq j$ and $a_{i,j}$ is any element of R.

In type 1, all the off-diagonal elements are zero. In type 2, there are two non-zero off-diagonal elements. In type 3, there is at most one non-zero off-diagonal element, and it may be above or below the diagonal.

Exercise Show that if B is an elementary matrix of type 1,2, or 3, then B is invertible and B^{-1} is an elementary matrix of the same type.

The following theorem is handy when working with matrices.

Theorem Suppose A is a matrix. It need not be square. To perform an elementary row (column) operation on A, perform the operation on an identity matrix to obtain an elementary matrix B, and multiply on the left (right). That is, BA = row operation on A and AB = column operation on A. (See the exercise on page 54.)

Exercise Suppose F is a field and $A \in F_{m,n}$.

1) Show $\exists$ invertible matrices $B \in F_m$ and $C \in F_n$ such that $BAC = (d_{i,j})$ where $d_{1,1} = \cdots = d_{t,t} = \underline{1}$ and all other entries are $\underline{0}$. The integer t is called the *rank* of A. (See page 89 of Chapter 5.)

2) Suppose $A \in F_n$ is invertible. Show A is the product of elementary matrices.

3) A matrix T is said to be in *row echelon* form if, for each $1 \le i < m$, the first non-zero term of row $(i+1)$ is to the right of the first non-zero term of row i. Show $\exists$ an invertible matrix $B \in F_m$ such that BA is in row echelon form.

4) Let $A = \begin{pmatrix} 3 & 11 \\ 0 & 4 \end{pmatrix}$ and $D = \begin{pmatrix} 3 & 11 \\ 1 & 4 \end{pmatrix}$. Write A and D as products of elementary matrices over $\mathbf{Q}$. Is it possible to write them as products of elementary matrices over $\mathbf{Z}$?

For 1), perform row and column operations on A to reach the desired form. This shows the matrices B and C may be selected as products of elementary matrices. Part 2) also follows from this procedure. For part 3), use only row operations. Notice that if T is in row-echelon form, the number of non-zero rows is the rank of T.

Systems of Equations

Suppose $A = (a_{i,j}) \in R_{m,n}$ and $C = \begin{pmatrix} c_1 \\ \cdot \\ \cdot \\ c_m \end{pmatrix} \in R^m = R_{m,1}$. The system

$$\begin{matrix} a_{1,1}x_1 + \cdots + a_{1,n}x_n = & c_1 \\ \vdots \qquad\qquad \vdots & \vdots \\ a_{m,1}x_1 + \cdots + a_{m,n}x_n = & c_m \end{matrix}$$

of m equations in n unknowns, can be written as one matrix equation in one unknown, namely as $(a_{i,j})\begin{pmatrix} x_1 \\ \cdot \\ \cdot \\ x_n \end{pmatrix} = \begin{pmatrix} c_1 \\ \cdot \\ \cdot \\ c_m \end{pmatrix}$ or $AX = C$.

Define $f : R^n \to R^m$ by $f(D) = AD$. Then f is a group homomorphism and also $f(Dc) = f(D)c$ for any $c \in R$. In the language of the next chapter, this says that f is an R-module homomorphism. The next theorem summarizes what we already know about solutions of linear equations in this setting.

Theorem

1) $AX = \underline{0}$ is called the *homogeneous equation.* Its solution set is $\ker(f)$.

2) $AX = C$ has a solution iff $C \in \text{image}(f)$. If $D \in R^n$ is one solution, the solution set $f^{-1}(C)$ is the coset $D + \ker(f)$ in R^n. (See part 7 of the theorem on homomorphisms in Chapter 2, page 28.)

3) Suppose $B \in R_m$ is invertible. Then $AX = C$ and $(BA)X = BC$ have the same set of solutions. Thus we may perform any row operation on both sides of the equation and not change the solution set.

4) If $m = n$ and $A \in R_m$ is invertible, then $AX = C$ has the unique solution $X = A^{-1}C$.

The geometry of systems of equations over a field will not become really transparent until the development of linear algebra in Chapter 5.

Determinants

The concept of determinant is one of the most amazing in all of mathematics. The proper development of this concept requires a study of multilinear forms, which is given in Chapter 6. In this section we simply present the basic properties.

For each $n \geq 1$ and each commutative ring R, determinant is a function from R_n to R. For $n = 1$, $|\,(a)\,| = a$. For $n = 2$, $\left| \begin{pmatrix} a & b \\ c & d \end{pmatrix} \right| = ad - bc$.

Definition Let $A = (a_{i,j}) \in R_n$. If σ is a permutation on $\{1, 2, ..., n\}$, let $\text{sign}(\sigma) = 1$ if σ is an even permutation, and $\text{sign}(\sigma) = -1$ if σ is an odd permutation. The *determinant* is defined by $|\,A\,| = \sum_{\text{all}\,\sigma} \text{sign}(\sigma)\, a_{1,\sigma(1)} \cdot a_{2,\sigma(2)} \cdots a_{n,\sigma(n)}$. Check that for $n = 2$, this agrees with the definition above. (Note that here we are writing the permutation functions as $\sigma(i)$ and not as $(i)\sigma$.)

For each σ, $a_{1,\sigma(1)} \cdot a_{2,\sigma(2)} \cdots a_{n,\sigma(n)}$ contains exactly one factor from each row and one factor from each column. Since R is commutative, we may rearrange the factors so that the first comes from the first column, the second from the second column, etc. This means that there is a permutation τ on $\{1, 2, \ldots, n\}$ such that $a_{1,\sigma(1)} \cdots a_{n,\sigma(n)} = a_{\tau(1),1} \cdots a_{\tau(n),n}$. We wish to show that $\tau = \sigma^{-1}$ and thus $\text{sign}(\sigma) = \text{sign}(\tau)$. To reduce the abstraction, suppose $\sigma(2) = 5$. Then the first expression will contain the factor $a_{2,5}$. In the second expression, it will appear as $a_{\tau(5),5}$, and so $\tau(5) = 2$. Anyway, τ is the inverse of σ and thus there are two ways to define determinant. It follows that the determinant of a matrix is equal to the determinant of its transpose.

Theorem $|A| = \sum\limits_{\text{all } \sigma} \text{sign}(\sigma) a_{1,\sigma(1)} \cdot a_{2,\sigma(2)} \cdots a_{n,\sigma(n)} = \sum\limits_{\text{all } \tau} \text{sign}(\tau) a_{\tau(1),1} \cdot a_{\tau(2),2} \cdots a_{\tau(n),n}$.

Corollary $|A| = |A^t|$.

You may view an $n \times n$ matrix A as a sequence of n column vectors or as a sequence of n row vectors. Here we will use column vectors. This means we write the matrix A as $A = (A_1, A_2, \ldots, A_n)$ where each $A_i \in R_{n,1} = R^n$.

Theorem If two columns of A are equal, then $|A| = \underline{0}$.

Proof For simplicity, assume the first two columns are equal, i.e., $A_1 = A_2$. Now $|A| = \sum\limits_{\text{all } \tau} \text{sign}(\tau) a_{\tau(1),1} \cdot a_{\tau(2),2} \cdots a_{\tau(n),n}$ and this summation has $n!$ terms and $n!$ is an even number. Let γ be the transposition which interchanges one and two. Then for any τ, $a_{\tau(1),1} \cdot a_{\tau(2),2} \cdots a_{\tau(n),n} = a_{\tau\gamma(1),1} \cdot a_{\tau\gamma(2),2} \cdots a_{\tau\gamma(n),n}$. This pairs up the $n!$ terms of the summation, and since $\text{sign}(\tau) = -\text{sign}(\tau\gamma)$, these pairs cancel in the summation. Therefore $|A| = \underline{0}$.

Theorem Suppose $1 \leq r \leq n$, $C_r \in R_{n,1}$, and $a, c \in R$. Then $|(A_1, \ldots, A_{r-1}, aA_r + cC_r, A_{r+1}, \ldots, A_n)| = a|(A_1, \ldots, A_n)| + c|(A_1, \ldots, A_{r-1}, C_r, A_{r+1}, \ldots, A_n)|$

Proof This is immediate from the definition of determinant and the distributive law of multiplication in the ring R.

Summary Determinant is a function $d : R_n \to R$. In the language used in the Appendix, the two previous theorems say that d is an alternating multilinear form. The next two theorems show that alternating implies skew-symmetric (see page 129).

Theorem Interchanging two columns of A multiplies the determinant by minus one.

Proof For simplicity, show that $|(A_2, A_1, A_3, \ldots, A_n)| = -|A|$. We know $\underline{0} = |(A_1 + A_2, A_1 + A_2, A_3, \ldots, A_n)| = |(A_1, A_1, A_3, \ldots, A_n)| + |(A_1, A_2, A_3, \ldots, A_n)| + |(A_2, A_1, A_3, \ldots, A_n)| + |(A_2, A_2, A_3, \ldots, A_n)|$. Since the first and last of these four terms are zero, the result follows.

Theorem If τ is a permutation of $(1, 2, \ldots, n)$, then
$|A| = \text{sign}(\tau)|(A_{\tau(1)}, A_{\tau(2)}, \ldots, A_{\tau(n)})|$.

Proof The permutation τ is the finite product of transpositions.

Exercise Rewrite the four preceding theorems using rows instead of columns.

The following theorem is just a summary of some of the work done so far.

Theorem Multiplying any row or column of matrix by a scalar $c \in R$, multiplies the determinant by c. Interchanging two rows or two columns multiplies the determinant by -1. Adding c times one row to another row, or adding c times one column to another column, does not change the determinant. If a matrix has two rows equal or two columns equal, its determinant is zero. More generally, if one row is c times another row, or one column is c times another column, then the determinant is zero.

There are $2n$ ways to compute $|A|$; expansion by any row or expansion by any column. Let $M_{i,j}$ be the determinant of the $(n-1) \times (n-1)$ matrix obtained by removing row i and column j from A. Let $C_{i,j} = (-1)^{i+j} M_{i,j}$. $M_{i,j}$ and $C_{i,j}$ are called the (i,j) *minor* and *cofactor* of A. The following theorem is useful but the proof is a little tedious and should not be done as an exercise.

Theorem For any $1 \leq i \leq n$, $|A| = a_{i,1}C_{i,1} + a_{i,2}C_{i,2} + \cdots + a_{i,n}C_{i,n}$. For any $1 \leq j \leq n$, $|A| = a_{1,j}C_{1,j} + a_{2,j}C_{2,j} + \cdots + a_{n,j}C_{n,j}$. Thus if any row or any column is zero, the determinant is zero.

Exercise Let $A = \begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{pmatrix}$. The determinant of A is the sum of six terms.

Write out the determinant of A expanding by the first column and also expanding by the second row.

Theorem If A is an upper or lower triangular matrix, $|A|$ is the product of the diagonal elements. If A is an elementary matrix of type 2, $|A| = -1$. If A is an elementary matrix of type 3, $|A| = 1$.

Proof We will prove the first statement for upper triangular matrices. If $A \in R_2$ is an upper triangular matrix, then its determinant is the product of the diagonal elements. Suppose $n > 2$ and the theorem is true for matrices in R_{n-1}. Suppose $A \in R_n$ is upper triangular. The result follows by expanding by the first column.

An elementary matrix of type 3 is a special type of upper or lower triangular matrix, so its determinant is 1. An elementary matrix of type 2 is obtained from the identity matrix by interchanging two rows or columns, and thus has determinant -1.

Theorem (Determinant by blocks) Suppose $A \in R_n$, $B \in R_{n,m}$, and $D \in R_m$. Then the determinant of $\begin{pmatrix} A & B \\ O & D \end{pmatrix}$ is $|A||D|$.

Proof Expand by the first column and use induction on n.

The following remarkable theorem takes some work to prove. We assume it here without proof. (For the proof, see page 130 of the Appendix.)

Theorem The determinant of the product is the product of the determinants, i.e., if $A, B \in R_n$, $|AB| = |A||B|$. Thus $|AB| = |BA|$ and if C is invertible, $|C^{-1}AC| = |ACC^{-1}| = |A|$.

Corollary If A is a unit in R_n, then $|A|$ is a unit in R and $|A^{-1}| = |A|^{-1}$.

Proof $\underline{1} = |I| = |AA^{-1}| = |A||A^{-1}|$.

One of the major goals of this chapter is to prove the converse of the preceding corollary.

Classical adjoint Suppose R is a commutative ring and $A \in R_n$. The *classical adjoint* of A is $(C_{i,j})^t$, i.e., the matrix whose (j, i) term is the (i, j) cofactor. Before

we consider the general case, let's examine 2×2 matrices.

If $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ then $(C_{i,j}) = \begin{pmatrix} d & -c \\ -b & a \end{pmatrix}$ and so $(C_{i,j})^t = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$. Then $A(C_{i,j})^t = (C_{i,j})^t A = \begin{pmatrix} |A| & 0 \\ 0 & |A| \end{pmatrix} = |A|\ I$. Thus if $|A|$ is a unit in R, A is invertible and $A^{-1} = |A|^{-1}\ (C_{i,j})^t$. In particular, if $|A| = \underline{1}$, $A^{-1} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$. Here is the general case.

Theorem If R is commutative and $A \in R_n$, then $A(C_{i,j})^t = (C_{i,j})^t A = |A|\ I$.

Proof We must show that the diagonal elements of the product $A(C_{i,j})^t$ are all $|A|$ and the other elements are $\underline{0}$. The (s, s) term is the dot product of row s of A with row s of $(C_{i,j})$ and is thus $|A|$ (computed by expansion by row s). For $s \neq t$, the (s, t) term is the dot product of row s of A with row t of $(C_{i,j})$. Since this is the determinant of a matrix with row s = row t, the (s, t) term is $\underline{0}$. The proof that $(C_{i,j})^t A = |A| I$ is similar and is left as an exercise.

We are now ready for one of the most beautiful and useful theorems in all of mathematics.

Theorem Suppose R is a commutative ring and $A \in R_n$. Then A is a unit in R_n iff $|A|$ is a unit in R. (Thus if R is a field, A is invertible iff $|A| \neq \underline{0}$.) If A is invertible, then $A^{-1} = |A|^{-1}\ (C_{i,j})^t$. Thus if $|A| = \underline{1}$, $A^{-1} = (C_{i,j})^t$, the classical adjoint of A.

Proof This follows immediately from the preceding theorem.

Exercise Show that any right inverse of A is also a left inverse. That is, suppose $A, B \in R_n$ and $AB = I$. Show A is invertible with $A^{-1} = B$, and thus $BA = I$.

Similarity

Suppose $A, B \in R_n$. B is said to be *similar* to A if $\exists$ an invertible $C \in R_n$ such that $B = C^{-1}AC$, i.e., B is similar to A iff B is a *conjugate* of A.

Theorem B is similar to B.

B is similar to A iff A is similar to B.

If D is similar to B and B is similar to A, then D is similar to A.

"Similarity" is an equivalence relation on R_n.

Proof This is a good exercise using the definition.

Theorem Suppose A and B are similar. Then $|A| = |B|$ and thus A is invertible iff B is invertible.

Proof Suppose $B = C^{-1}AC$. Then $|B| = |C^{-1}AC| = |ACC^{-1}| = |A|$.

Trace Suppose $A = (a_{i,j}) \in R_n$. Then the *trace* is defined by trace$(A) = a_{1,1} + a_{2,2} + \cdots + a_{n,n}$. That is, the trace of A is the sum of its diagonal terms.

One of the most useful properties of trace is trace$(AB) = $ trace(BA) whenever AB and BA are defined. For example, suppose $A = (a_1, a_2, ..., a_n)$ and $B = (b_1, b_2, ..., b_n)^t$. Then AB is the scalar $a_1b_1 + \cdots + a_nb_n$ while BA is the $n \times n$ matrix (b_ia_j). Note that trace$(AB) = $ trace(BA). Here is the theorem in full generality.

Theorem Suppose $A \in R_{m,n}$ and $B \in R_{n,m}$. Then AB and BA are square matrices with trace$(AB) = $ trace(BA).

Proof This proof involves a change in the order of summation. By definition, $\text{trace}(AB) = \sum\limits_{1\le i\le m} a_{i,1}b_{1,i} + \cdots + a_{i,n}b_{n,i} = \sum\limits_{\substack{1\le i\le m\\ 1\le j\le n}} a_{i,j}b_{j,i} = \sum\limits_{1\le j\le n} b_{j,1}a_{1,j} + \cdots + b_{j,m}a_{m,j} = \text{trace}(BA)$.

Theorem If $A, B \in R_n$, trace$(A + B) = $ trace$(A) + $ trace(B) and trace$(AB) = $ trace(BA).

Proof The first part of the theorem is immediate, and the second part is a special case of the previous theorem.

Theorem If A and B are similar, then trace$(A) = $ trace(B).

Proof trace$(B) = $ trace$(C^{-1}AC) = $ trace$(ACC^{-1}) = $ trace(A).

Summary Determinant and trace are functions from R_n to R. Determinant is a multiplicative homomorphism and trace is an additive homomorphism. Furthermore $|AB| = |BA|$ and $\text{trace}(AB) = \text{trace}(BA)$. If A and B are similar, $|A| = |B|$ and $\text{trace}(A) = \text{trace}(B)$.

Exercise Suppose $A \in R_n$ and $a \in R$. Find $|aA|$ and $\text{trace}(aA)$.

Characteristic polynomials If $A \in R_n$, the *characteristic polynomial* $CP_A(x) \in R[x]$ is defined by $CP_A(x) = |(xI - A)|$. Any $\lambda \in R$ which is a root of $CP_A(x)$ is called a *characteristic root* of A.

Theorem $CP_A(x) = a_0 + a_1x + \cdots + a_{n-1}x^{n-1} + x^n$ where $\text{trace}(A) = -a_{n-1}$ and $|A| = (-1)^n a_0$.

Proof This follows from a direct computation of the determinant.

Theorem If A and B are similar, then they have the same characteristic polynomials.

Proof Suppose $B = C^{-1}AC$. $CP_B(x) = |(xI - C^{-1}AC)| = |C^{-1}(xI - A)C| = |(xI - A)| = CP_A(x)$.

Exercise Suppose R is a commutative ring, $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is a matrix in R_2, and $CP_A(x) = a_0 + a_1x + x^2$. Find a_0 and a_1 and show that $a_0I + a_1A + A^2 = \underline{0}$, i.e., show A satisfies its characteristic polynomial. In other words, $CP_A(A) = \underline{0}$.

Exercise Suppose F is a field and $A \in F_2$. Show the following are equivalent.

1) $A^2 = \underline{0}$.
2) $|A| = \text{trace}(A) = \underline{0}$.
3) $CP_A(x) = x^2$.
4) $\exists$ an elementary matrix C such that $C^{-1}AC$ is strictly upper triangular.

Note This exercise is a special case of a more general theorem. A square matrix over a field is nilpotent iff all its characteristic roots are $\underline{0}$ iff it is similar to a strictly upper triangular matrix. This remarkable result cannot be proved by matrix theory alone, but depends on linear algebra (see pages 93, 94, and 98).

Chapter 5

Linear Algebra

The exalted position held by linear algebra is based upon the subject's ubiquitous utility and ease of application. The basic theory is developed here in full generality, i.e., modules are defined over an arbitrary ring R and not just over a field. The elementary facts about cosets, quotients, and homomorphisms follow the same pattern as in the chapters on groups and rings. We give a simple proof that if R is a commutative ring and $f : R^n \to R^n$ is a surjective R-module homomorphism, then f is an isomorphism. This shows that finitely generated free R-modules have a well defined dimension, and simplifies some of the development of linear algebra. It is in this chapter that the concepts about functions, solutions of equations, matrices, and generating sets come together in one unified theory.

After the general theory, we restrict our attention to vector spaces, i.e., modules over a field. The key theorem is that any vector space V has a free basis, and thus if V is finitely generated, it has a well defined dimension, and incredible as it may seem, this single integer determines V up to isomorphism. Also any endomorphism $f : V \to V$ may be represented by a matrix, and any change of basis corresponds to conjugation of that matrix. One of the goals in linear algebra is to select a basis so that the matrix representing f has a simple form. For example, if f is not injective, then f may be represented by a matrix whose first column is zero. As another example, if f is nilpotent, then f may be represented by a strictly upper triangular matrix. The theorem on Jordan canonical form is not proved in this chapter, and should not be considered part of this chapter. It is stated here in full generality only for reference and completeness. The proof is given in the Appendix. This chapter concludes with the study of real inner product spaces, and with the beautiful theory relating orthogonal matrices and symmetric matrices.

Definition Suppose R is a ring and M is an additive abelian group. The statement that M is a *right R-module* means there is a scalar multiplication

$$\begin{array}{lll} M \times R \to M & \text{satisfying} & (a_1 + a_2)r = a_1r + a_2r \\ (m, r) \to mr & & a(r_1 + r_2) = ar_1 + ar_2 \\ & & a(r_1 \cdot r_2) = (ar_1)r_2 \\ & & a\underline{1} = a \end{array}$$

for all $a, a_1, a_2 \in M$ and $r, r_1, r_2 \in R$.

The statement that M is a *left R-module* means there is a scalar multiplication

$$\begin{array}{lll} R \times M \to M & \text{satisfying} & r(a_1 + a_2) = ra_1 + ra_2 \\ (r, m) \to rm & & (r_1 + r_2)a = r_1a + r_2a \\ & & (r_1 \cdot r_2)a = r_1(r_2a) \\ & & \underline{1}a = a \end{array}$$

Note that the plus sign is used ambiguously, as addition in M and as addition in R.

Notation The fact that M is a right (left) R-module will be denoted by $M = M_R$ ($M = {}_RM$). If R is commutative and $M = M_R$ then left scalar multiplication defined by $ra = ar$ makes M into a left R-module. Thus for commutative rings, we may write the scalars on either side. In this text we stick to right R-modules.

Convention Unless otherwise stated, it is assumed that R is a ring and the word "R-module" (or sometimes just "module") means "right R-module".

Theorem Suppose M is an R-module.

1) If $r \in R$, then $f : M \to M$ defined by $f(a) = ar$ is a homomorphism of additive groups. In particular $(\underline{0}_M)r = \underline{0}_M$.

2) If $a \in M$, $a\underline{0}_R = \underline{0}_M$.

3) If $a \in M$ and $r \in R$, then $(-a)r = -(ar) = a(-r)$.

Proof This is a good exercise in using the axioms for an R-module.

Submodules If M is an R-module, the statement that a subset $N \subset M$ is a *submodule* means it is a subgroup which is closed under scalar multiplication, i.e., if $a \in N$ and $r \in R$, then $ar \in N$. In this case N will be an R-module because the axioms will automatically be satisfied. Note that $\underline{0}$ and M are submodules, called the *improper* submodules of M.

Theorem Suppose M is an R-module, T is an index set, and for each $t \in T$, N_t is a submodule of M.

1) $\bigcap_{t \in T} N_t$ is a submodule of M.
2) If $\{N_t\}$ is a monotonic collection, $\bigcup_{t \in T} N_t$ is a submodule.
3) $+_{t \in T} N_t = \{$all finite sums $a_1 + \cdot\cdot + a_m$: each a_i belongs to some $N_t\}$ is a submodule. If $T = \{1, 2, .., n\}$, then this submodule may be written as $N_1 + N_2 + \cdot\cdot + N_n = \{a_1 + a_2 + \cdot\cdot + a_n : \text{ each } a_i \in N_i\}$.

Proof We know from page 22 that versions of 1) and 2) hold for subgroups, and in particular for subgroups of additive abelian groups. To finish the proofs it is only necessary to check scalar multiplication, which is immediate. Also the proof of 3) is immediate. Note that if N_1 and N_2 are submodules of M, $N_1 + N_2$ is the smallest submodule of M containing $N_1 \cup N_2$.

Exercise Suppose T is a non-void set, N is an R-module, and N^T is the collection of all functions $f : T \to N$ with addition defined by $(f+g)(t) = f(t)+g(t)$, and scalar multiplication defined by $(fr)(t) = f(t)r$. Show N^T is an R-module. (We know from the last exercise in Chapter 2 that N^T is a group, and so it is only necessary to check scalar multiplication.) This simple fact is quite useful in linear algebra. For example, in 5) of the theorem below, it is stated that $\text{Hom}_R(M, N)$ forms an abelian group. So it is only necessary to show that $\text{Hom}_R(M, N)$ is a subgroup of N^M. Also in 8) it is only necessary to show that $\text{Hom}_R(M, N)$ is a submodule of N^M.

Homomorphisms

Suppose M and N are R-modules. A function $f : M \to N$ is a *homomorphism* (i.e., an R-module homomorphism) provided it is a group homomorphism and if $a \in M$ and $r \in R$, $f(ar) = f(a)r$. On the left, scalar multiplication is in M and on the right it is in N. The basic facts about homomorphisms are listed below. Much

of this work has already been done in the chapter on groups (see page 28).

Theorem

1) The zero map $M \to N$ is a homomorphism.

2) The identity map $I : M \to M$ is a homomorphism.

3) The composition of homomorphisms is a homomorphism.

4) The sum of homomorphisms is a homomorphism. If $f, g : M \to N$ are homomorphisms, define $(f + g) : M \to N$ by $(f + g)(a) = f(a) + g(a)$. Then $f + g$ is a homomorphism. Also $(-f)$ defined by $(-f)(a) = -f(a)$ is a homomorphism. If $h : N \to P$ is a homomorphism, $h \circ (f + g) = (h \circ f) + (h \circ g)$. If $k : P \to M$ is a homomorphism, $(f + g) \circ k = (f \circ k) + (g \circ k)$.

5) $\text{Hom}_R(M, N) = \text{Hom}(M_R, N_R)$, the set of all homomorphisms from M to N, forms an abelian group under addition. $\text{Hom}_R(M, M)$, with multiplication defined to be composition, is a ring.

6) If a bijection $f : M \to N$ is a homomorphism, then $f^{-1} : N \to M$ is also a homomorphism. In this case f and f^{-1} are called *isomorphisms.* A homomorphism $f : M \to M$ is called an *endomorphism.* An isomorphism $f : M \to M$ is called an *automorphism.* The units of the endomorphism ring $\text{Hom}_R(M, M)$ are the automorphisms. Thus the automorphisms on M form a group under composition. We will see later that if $M = R^n$, $\text{Hom}_R(R^n, R^n)$ is just the matrix ring R_n and the automorphisms are merely the invertible matrices.

7) If R is commutative and $r \in R$, then $g : M \to M$ defined by $g(a) = ar$ is a homomorphism. Furthermore, if $f : M \to N$ is a homomorphism, fr defined by $(fr)(a) = f(ar) = f(a)r$ is a homomorphism.

8) If R is commutative, $\text{Hom}_R(M, N)$ is an R-module.

9) Suppose $f : M \to N$ is a homomorphism, $G \subset M$ is a submodule, and $H \subset N$ is a submodule. Then $f(G)$ is a submodule of N and $f^{-1}(H)$ is a submodule of M. In particular, image(f) is a submodule of N and $\ker(f) = f^{-1}(\underline{0})$ is a submodule of M.

Proof This is just a series of observations.

Abelian groups are Z-modules On page 21, it is shown that any additive group M admits a scalar multiplication by integers, and if M is abelian, the properties are satisfied to make M a **Z**-module. Note that this is the only way M can be a **Z**-module, because $a1 = a$, $a2 = a + a$, etc. Furthermore, if $f : M \to N$ is a group homomorphism of abelian groups, then f is also a **Z**-module homomorphism.

Summary Additive abelian groups are "the same things" as **Z**-modules. While group theory in general is quite separate from linear algebra, the study of additive abelian groups is a special case of the study of R-modules.

Exercise R-modules are also **Z**-modules and R-module homomorphisms are also **Z**-module homomorphisms. If M and N are **Q**-modules and $f : M \to N$ is a **Z**-module homomorphism, must it also be a **Q**-module homomorphism?

Homomorphisms on R^n

R^n **as an** R**-module** On page 54 it was shown that the additive abelian group $R_{m,n}$ admits a scalar multiplication by elements in R. The properties listed there were exactly those needed to make $R_{m,n}$ an R-module. Of particular importance is the case $R^n = R \oplus \cdot\cdot \oplus R = R_{n,1}$ (see page 53). We begin with the case $n = 1$.

R **as a right** R**-module** Let $M = R$ and define scalar multiplication on the right by $ar = a \cdot r$. That is, scalar multiplication is just ring multiplication. This makes R a right R-module denoted by R_R (or just R). This is the same as the definition before for R^n when $n = 1$.

Theorem Suppose R is a ring and N is a subset of R. Then N is a submodule of R_R $({}_RR)$ iff N is a right (left) ideal of R.

Proof The definitions are the same except expressed in different language.

Theorem Suppose $M = M_R$ and $f, g : R \to M$ are homomorphisms with $f(\underline{1}) = g(\underline{1})$. Then $f = g$. Furthermore, if $m \in M$, $\exists!$ homomorphism $h : R \to M$ with $h(\underline{1}) = m$. In other words, $\text{Hom}_R(R, M) \approx M$.

Proof Suppose $f(\underline{1}) = g(\underline{1})$. Then $f(r) = f(\underline{1} \cdot r) = f(\underline{1})r = g(\underline{1})r = g(\underline{1} \cdot r) = g(r)$. Given $m \in M$, $h : R \to M$ defined by $h(r) = mr$ is a homomorphism. Thus

evaluation at $\underline{1}$ gives a bijection from $\text{Hom}_R(R, M)$ to M, and this bijection is clearly a group isomorphism. If R is commutative, it is an isomorphism of R-modules.

In the case $M = R$, the above theorem states that multiplication on left by some $m \in R$ defines a right R-module homomorphism from R to R, and every module homomorphism is of this form. The element m should be thought of as a 1×1 matrix. We now consider the case where the domain is R^n.

Homomorphisms on R^n Define $e_i \in R^n$ by $e_i = \begin{pmatrix} \underline{0} \\ \cdot \\ \underline{1}_i \\ \cdot \\ \underline{0} \end{pmatrix}$. Note that any $\begin{pmatrix} r_1 \\ \cdot \\ \\ \cdot \\ r_n \end{pmatrix}$ can be written uniquely as $e_1r_1 + \cdot\cdot + e_nr_n$. The sequence $\{e_1, .., e_n\}$ is called the *canonical free basis* or *standard basis* for R^n.

Theorem Suppose $M = M_R$ and $f, g : R^n \to M$ are homomorphisms with $f(e_i) = g(e_i)$ for $1 \leq i \leq n$. Then $f = g$. Furthermore, if $m_1, m_2, ..., m_n \in M$, $\exists!$ homomorphism $h : R^n \to M$ with $h(e_i) = m_i$ for $1 \leq i \leq m$. The homomorphism h is defined by $h(e_1r_1 + \cdot\cdot + e_nr_n) = m_1r_1 + \cdot\cdot + m_nr_n$.

Proof The proof is straightforward. Note this theorem gives a bijection from $\text{Hom}_R(R^n, M)$ to $M^n = M \times M \times \cdot\cdot \times M$ and this bijection is a group isomorphism. We will see later that the product M^n is an R-module with scalar multiplication defined by $(m_1, m_2, .., m_n)r = (m_1r, m_2r, .., m_nr)$. If R is commutative so that $\text{Hom}_R(R^n, M)$ is an R-module, this theorem gives an R-module isomorphism from $\text{Hom}_R(R^n, M)$ to M^n.

This theorem reveals some of the great simplicity of linear algebra. It does not matter how complicated the ring R is, or which R-module M is selected. Any R-module homomorphism from R^n to M is determined by its values on the basis, and any function from that basis to M extends uniquely to a homomorphism from R^n to M.

Exercise Suppose R is a field and $f : R_R \to M$ is a non-zero homomorphism. Show f is injective.

Now let's examine the special case $M = R^m$ and show $\text{Hom}_R(R^n, R^m) \approx R_{m,n}$.

Theorem Suppose $A = (a_{i,j}) \in R_{m,n}$. Then $f : R^n \to R^m$ defined by $f(B) = AB$ is a homomorphism with $f(e_i) =$ column i of A. Conversely, if $v_1, \ldots, v_n \in R^m$, define $A \in R_{m,n}$ to be the matrix with column $i = v_i$. Then f defined by $f(B) = AB$ is the unique homomorphism from R^n to R^m with $f(e_i) = v_i$.

Even though this follows easily from the previous theorem and properties of matrices, it is one of the great classical facts of linear algebra. Matrices over R give R-module homomorphisms! Furthermore, addition of matrices corresponds to addition of homomorphisms, and multiplication of matrices corresponds to composition of homomorphisms. These properties are made explicit in the next two theorems.

Theorem If $f, g : R^n \to R^m$ are given by matrices $A, C \in R_{m,n}$, then $f + g$ is given by the matrix $A+C$. Thus $\text{Hom}_R(R^n, R^m)$ and $R_{m,n}$ are isomorphic as additive groups. If R is commutative, they are isomorphic as R-modules.

Theorem If $f\colon R^n \to R^m$ is the homomorphism given by $A \in R_{m,n}$ and $g : R^m \to R^p$ is the homomorphism given by $C \in R_{p,m}$, then $g \circ f : R^n \to R^p$ is given by $CA \in R_{p,n}$. That is, composition of homomorphisms corresponds to multiplication of matrices.

Proof This is just the associative law of matrix multiplication, $C(AB) = (CA)B$.

The previous theorem reveals where matrix multiplication comes from. It is the matrix which represents the composition of the functions. In the case where the domain and range are the same, we have the following elegant corollary.

Corollary $\text{Hom}_R(R^n, R^n)$ and R_n are isomorphic as rings. The automorphisms correspond to the invertible matrices.

This corollary shows one way non-commutative rings arise, namely as endomorphism rings. Even if R is commutative, R_n is never commutative unless $n = 1$.

We now return to the general theory of modules (over some given ring R).

Cosets and Quotient Modules

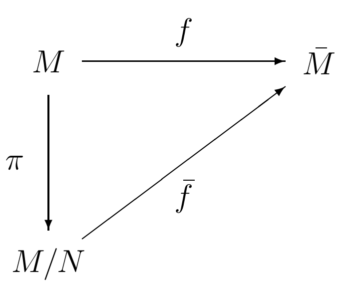
After seeing quotient groups and quotient rings, quotient modules go through without a hitch. As before, R is a ring and module means R-module.

Theorem Suppose M is a module and $N \subset M$ is a submodule. Since N is a normal subgroup of M, the additive abelian quotient group M/N is defined. Scalar multiplication defined by $(a+N)r = (ar+N)$ is well-defined and gives M/N the structure of an R-module. The natural projection $\pi : M \to M/N$ is a surjective homomorphism with kernel N. Furthermore, if $f : M \to \bar{M}$ is a surjective homomorphism with $\ker(f) = N$, then $M/N \approx \bar{M}$ (see below).

Proof On the group level, this is all known from Chapter 2 (see pages 27 and 29). It is only necessary to check the scalar multiplication, which is obvious.

The relationship between quotients and homomorphisms for modules is the same as for groups and rings, as shown by the next theorem.

Theorem Suppose $f : M \to \bar{M}$ is a homomorphism and N is a submodule of M. If $N \subset \ker(f)$, then $\bar{f} : (M/N) \to \bar{M}$ defined by $\bar{f}(a+N) = f(a)$ is a well-defined homomorphism making the following diagram commute.



Thus defining a homomorphism on a quotient module is the same as defining a homomorphism on the numerator that sends the denominator to $\underline{0}$. The image of $\bar{f}$ is the image of f, and the kernel of $\bar{f}$ is $\ker(f)/N$. Thus if $N = \ker(f)$, $\bar{f}$ is injective, and thus $(M/N) \approx \text{image}(f)$. Therefore for any homomorphism f, $(\text{domain}(f)/\ker(f)) \approx \text{image}(f)$.

Proof On the group level this is all known from Chapter 2 (see page 29). It is only necessary to check that $\bar{f}$ is a module homomorphism, and this is immediate.

Theorem Suppose M is an R-module and K and L are submodules of M.

i) The natural homomorphism $K \to (K+L)/L$ is surjective with kernel $K \cap L$. Thus $(K/K \cap L) \xrightarrow{\approx} (K+L)/L$ is an isomorphism.
ii) Suppose $K \subset L$. The natural homomorphism $M/K \to M/L$ is surjective with kernel L/K. Thus $(M/K)/(L/K) \xrightarrow{\approx} M/L$ is an isomorphism.

Examples These two examples are for the case $R = \mathbf{Z}$, i.e., for abelian groups.

1) $M = \mathbf{Z}$ $\quad K = 3\mathbf{Z}$ $\quad L = 5\mathbf{Z}$ $\quad K \cap L = 15\mathbf{Z}$ $\quad K + L = \mathbf{Z}$
$K/K \cap L = 3\mathbf{Z}/15\mathbf{Z} \approx \mathbf{Z}/5\mathbf{Z} = (K+L)/L$

2) $M = \mathbf{Z}$ $\quad K = 6\mathbf{Z}$ $\quad L = 3\mathbf{Z}$ $\quad (K \subset L)$
$(M/K)/(L/K) = (\mathbf{Z}/6\mathbf{Z})/(3\mathbf{Z}/6\mathbf{Z}) \approx \mathbf{Z}/3\mathbf{Z} = M/L$

Products and Coproducts

Infinite products work fine for modules, just as they do for groups and rings. This is stated below in full generality, although the student should think of the finite case. In the finite case something important holds for modules that does not hold for non-abelian groups or rings – namely, the finite product is also a coproduct. This makes the structure of module homomorphisms much more simple. For the finite case we may use either the product or sum notation, i.e., $M_1 \times M_2 \times \cdot \cdot \times M_n = M_1 \oplus M_2 \oplus \cdot \cdot \oplus M_n$.

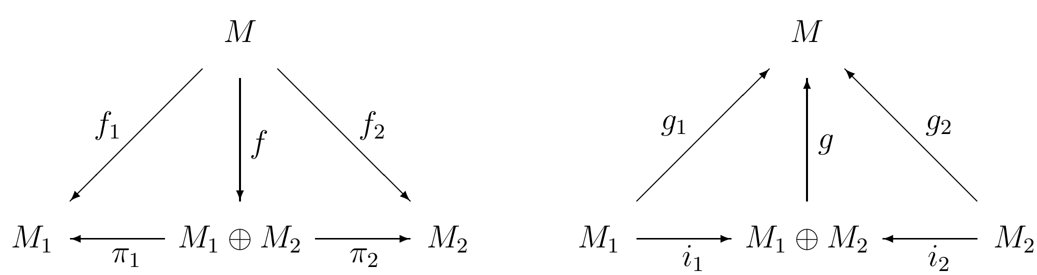
Theorem Suppose T is an index set and for each $t \in T$, M_t is an R-module. On the additive abelian group $\prod_{t \in T} M_t = \prod M_t$ define scalar multiplication by $\{m_t\}r = \{m_t r\}$. Then $\prod M_t$ is an R-module and, for each $s \in T$, the natural projection $\pi_s : \prod M_t \to M_s$ is a homomorphism. Suppose M is a module. Under the natural 1-1 correspondence from $\{$functions $f : M \to \prod M_t\}$ to $\{$sequence of functions $\{f_t\}_{t \in T}$ where $f_t : M \to M_t\}$, f is a homomorphism iff each f_t is a homomorphism.

Proof We already know from Chapter 2 that f is a group homomorphism iff each f_t is a group homomorphism. Since scalar multiplication is defined coordinatewise, f is a module homomorphism iff each f_t is a module homomorphism.

Definition If T is finite, the coproduct and product are the same module. If T is infinite, the *coproduct* or *sum* $\coprod_{t\in T} M_t = \bigoplus_{t\in T} M_t = \oplus M_t$ is the submodule of $\prod M_t$ consisting of all sequences $\{m_t\}$ with only a finite number of non-zero terms. For each $s \in T$, the inclusion homomorphisms $i_s : M_s \to \oplus M_t$ is defined by $i_s(a) = \{a_t\}$ where $a_t = \underset{\sim}{0}$ if $t \neq s$ and $a_s = a$. Thus each M_s may be considered to be a submodule of $\oplus M_t$.

Theorem Suppose M is an R-module. There is a 1-1 correspondence from $\{$homomorphisms $g : \oplus M_t \to M\}$ and $\{$sequences of homomorphisms $\{g_t\}_{t\in T}$ where $g_t : M_t \to M\}$. Given g, g_t is defined by $g_t = g \circ i_t$. Given $\{g_t\}$, g is defined by $g(\{m_t\}) = \sum_t g_t(m_t)$. Since there are only a finite number of non-zero terms, this sum is well defined.

For $T = \{1, 2\}$ the product and sum properties are displayed in the following commutative diagrams.



Theorem For finite T, the 1-1 correspondences in the above theorems actually produce group isomorphisms. If R is commutative, they give isomorphisms of R-modules.

$$\mathrm{Hom}_R(M, M_1 \oplus \cdot\cdot \oplus M_n) \approx \mathrm{Hom}_R(M, M_1) \oplus \cdot\cdot \oplus \mathrm{Hom}_R(M, M_n) \quad \text{and}$$
$$\mathrm{Hom}_R(M_1 \oplus \cdot\cdot \oplus M_n, M) \approx \mathrm{Hom}_R(M_1, M) \oplus \cdot\cdot \oplus \mathrm{Hom}_R(M_n, M)$$

Proof Let's look at this theorem for products with $n = 2$. All it says is that if $f = (f_1, f_2)$ and $h = (h_1, h_2)$, then $f + h = (f_1 + h_1, f_2 + h_2)$. If R is commutative, so that the objects are R-modules and not merely additive groups, then the isomorphisms are module isomorphisms. This says merely that $fr = (f_1, f_2)r = (f_1 r, f_2 r)$.

Exercise Suppose M and N are R-modules. Show that $M \oplus N$ is isomorphic to $N \oplus M$. Now suppose $A \subset M$, $B \subset N$ are submodules and show $(M \oplus N)/(A \oplus B)$ is isomorphic to $(M/A) \oplus (N/B)$. In particular, if $a \in R$ and $b \in R$, then $(R \oplus R)/(aR \oplus bR)$ is isomorphic to $(R/aR) \oplus (R/bR)$. For example, the abelian group $(\mathbf{Z} \oplus \mathbf{Z})/(2\mathbf{Z} \oplus 3\mathbf{Z})$ is isomorphic to $\mathbf{Z}_2 \oplus \mathbf{Z}_3$. These isomorphisms are transparent and are used routinely in algebra without comment (see Th 4, page 118).

Exercise Suppose R is a commutative ring, M is an R-module, and $n \geq 1$. Define a function $\alpha : \text{Hom}_R(R^n, M) \to M^n$ which is a R-module isomorphism.

Summands

One basic question in algebra is "When does a module split as the sum of two modules?". Before defining summand, here are two theorems for background.

Theorem Consider $M_1 = M_1 \oplus \underline{0}$ as a submodule of $M_1 \oplus M_2$. Then the projection map $\pi_2 : M_1 \oplus M_2 \to M_2$ is a surjective homomorphism with kernel M_1. Thus $(M_1 \oplus M_2)/M_1$ is isomorphic to M_2. (See page 35 for the group version.)

This is exactly what you would expect, and the next theorem is almost as intuitive.

Theorem Suppose K and L are submodules of M and $f : K \oplus L \to M$ is the natural homomorphism, $f(k, l) = k + l$. Then the image of f is $K + L$ and the kernel of f is $\{(a, -a) : a \in K \cap L\}$. Thus f is an isomorphism iff $K + L = M$ and $K \cap L = \underline{0}$. In this case we write $K \oplus L = M$. This abuse of notation allows us to avoid talking about "internal" and "external" direct sums.

Definition Suppose K is a submodule of M. The statement that K is a *summand* of M means $\exists$ a submodule L of M with $K \oplus L = M$. According to the previous theorem, this is the same as there exists a submodule L with $K + L = M$ and $K \cap L = \underline{0}$. If such an L exists, it need not be unique, but it will be unique up to isomorphism, because $L \approx M/K$. Of course, M and $\underline{0}$ are always summands of M.

Exercise Suppose M is a module and $K = \{(m, m) : m \in M\} \subset M \oplus M$. Show K is a submodule of $M \oplus M$ which is a summand.

Exercise $\mathbf{R}$ is a module over $\mathbf{Q}$, and $\mathbf{Q} \subset \mathbf{R}$ is a submodule. Is $\mathbf{Q}$ a summand of $\mathbf{R}$? With the material at hand, this is not an easy question. Later on, it will be easy.

Exercise Answer the following questions about abelian groups, i.e., $\mathbf{Z}$-modules. (See the third exercise on page 35.)

1) Is $2\mathbf{Z}$ a summand of $\mathbf{Z}$?
2) Is $2\mathbf{Z}_4$ a summand of $\mathbf{Z}_4$?
3) Is $3\mathbf{Z}_{12}$ a summand of $\mathbf{Z}_{12}$?
4) Suppose $m, n > 1$. When is $n\mathbf{Z}_{mn}$ a summand of $\mathbf{Z}_{mn}$?

Exercise If T is a ring, define the center of T to be the subring $\{t : ts = st \text{ for all } s \in T\}$. Let R be a commutative ring and $T = R_n$. There is a exercise on page 57 to show that the center of T is the subring of scalar matrices. Show R^n is a left T-module and find $\text{Hom}_T(R^n, R^n)$.

Independence, Generating Sets, and Free Basis

This section is a generalization and abstraction of the brief section **Homomorphisms on** R^n. These concepts work fine for an infinite index set T because linear combination means finite linear combination. However, to avoid dizziness, the student should first consider the case where T is finite.

Definition Suppose M is an R-module, T is an index set, and for each $t \in T$, $s_t \in M$. Let S be the sequence $\{s_t\}_{t\in T} = \{s_t\}$. The statement that S is *dependent* means $\exists$ a finite number of distinct elements $t_1, ..., t_n$ in T, and elements $r_1, .., r_n$ in R, not all zero, such that the linear combination $s_{t_1}r_1 + \cdots + s_{t_n}r_n = \underline{0}$. Otherwise, S is *independent*. Note that if some $s_t = \underline{0}$, then S is dependent. Also if $\exists$ distinct elements t_1 and t_2 in T with $s_{t_1} = s_{t_2}$, then S is dependent.

Let SR be the set of all linear combinations $s_{t_1}r_1 + \cdots + s_{t_n}r_n$. SR is a submodule of M called the submodule *generated* by S. If S is independent and generates M, then S is said to be a *basis* or *free basis* for M. In this case any $v \in M$ can be written uniquely as a linear combination of elements in S. An R-module M is said to be a *free* R-module if it is zero or if it has a basis. The next two theorems are obvious, except for the confusing notation. You might try first the case $T = \{1, 2, ..., n\}$ and $\oplus R_t = R^n$ (see p 72).

Theorem For each $t \in T$, let $R_t = R_R$ and for each $c \in T$, let $e_c \in \oplus R_t = \bigoplus_{t\in T} R_t$ be $e_c = \{r_t\}$ where $r_c = \underline{1}$ and $r_t = \underline{0}$ if $t \neq c$. Then $\{e_c\}_{c\in T}$ is a basis for $\oplus R_t$ called the *canonical basis* or *standard basis.*

Theorem Suppose N is an R-module and M is a free R-module with a basis $\{s_t\}$. Then $\exists$ a 1-1 correspondence from the set of all functions $g:\{s_t\} \to N$ and the set of all homomorphisms $f: M \to N$. Given g, define f by $f(s_{t_1}r_1 + \cdot\cdot + s_{t_n}r_n) = g(s_{t_1})r_1 + \cdot\cdot + g(s_{t_n})r_n$. Given f, define g by $g(s_t) = f(s_t)$. In other words, f is completely determined by what it does on the basis S, and you are "free" to send the basis any place you wish and extend to a homomorphism.

Recall that we have already had the preceding theorem in the case S is the canonical basis for $M = R^n$ (p 72). The next theorem is so basic in linear algebra that it is used without comment. Although the proof is easy, it should be worked carefully.

Theorem Suppose N is a module, M is a free module with basis $S = \{s_t\}$, and $f : M \to N$ is a homomorphism. Let $f(S)$ be the sequence $\{f(s_t)\}$ in N.

1) $f(S)$ generates N iff f is surjective.
2) $f(S)$ is independent in N iff f is injective.
3) $f(S)$ is a basis for N iff f is an isomorphism.
4) If $h : M \to N$ is a homomorphism, then $f = h$ iff $f \mid S = h \mid S$.

Exercise Let $(A_1, .., A_n)$ be a sequence of n vectors with each $A_i \in \mathbf{Z}^n$. Show this sequence is linearly independent over $\mathbf{Z}$ iff it is linearly independent over $\mathbf{Q}$. Is it true the sequence is linearly independent over $\mathbf{Z}$ iff it is linearly independent over $\mathbf{R}$? This question is difficult until we learn more linear algebra.

Characterization of Free Modules

Any free R-module is isomorphic to one of the canonical free R-modules $\oplus R_t$. This is just an observation, but it is a central fact in linear algebra.

Theorem A non-zero R-module M is free iff $\exists$ an index set T such that $M \approx \bigoplus_{t \in T} R_t$. In particular, M has a finite free basis of n elements iff $M \approx R^n$.

Proof If M is isomorphic to $\oplus R_t$ then M is certainly free. So now suppose M has a free basis $\{s_t\}$. Then the homomorphism $f : M \to \oplus R_t$ with $f(s_t) = e_t$ sends the basis for M to the canonical basis for $\oplus R_t$. By 3) in the preceding theorem, f is an isomorphism.

Exercise Suppose R is a commutative ring, $A \in R_n$, and the homomorphism $f : R^n \to R^n$ defined by $f(B) = AB$ is surjective. Show f is an isomorphism, i.e., show A is invertible. This is a key theorem in linear algebra, although it is usually stated only for the case where R is a field. Use the fact that $\{e_1, .., e_n\}$ is a free basis for R^n.

The next exercise is routine, but still informative.

Exercise Let $R = \mathbf{Z}$, $A = \begin{pmatrix} 2 & 1 & 0 \\ 3 & 2 & -5 \end{pmatrix}$ and $f \colon \mathbf{Z}^3 \to \mathbf{Z}^2$ be the group homomorphism defined by A. Find a non-trivial linear combination of the columns of A which is $\underline{0}$. Also find a non-zero element of kernel(f).

If R is a commutative ring, you can relate properties of R as an R-module to properties of R as a ring.

Exercise Suppose R is a commutative ring and $v \in R$, $v \neq \underline{0}$.

1) v is independent iff v is ____________________.
2) v is a basis for R iff v generates R iff v is ____________________.

Note that 2) here is essentially the first exercise for the case $n = 1$. That is, if $f : R \to R$ is a surjective R-module homomorphism, then f is an isomorphism.

Relating these concepts to matrices

The theorem stated below gives a summary of results we have already had. It shows that certain concepts about matrices, linear independence, injective homomorphisms, and solutions of equations, are all the same — they are merely stated in different language. Suppose $A \in R_{m,n}$ and $f : R^n \to R^m$ is the homomorphism associated with A, i.e., $f(B) = AB$. Let $v_1, .., v_n \in R^m$ be the columns of A, i.e., $f(e_i) = v_i$ = column i of A. Let $B = \begin{pmatrix} b_1 \\ \cdot \\ b_n \end{pmatrix}$ represent an element of R^n and $C = \begin{pmatrix} c_1 \\ \cdot \\ c_m \end{pmatrix}$

represent an element of R^m.

Theorem

1) The element $f(B)$ is a linear combination of the columns of A, that is $f(B) = f(e_1b_1 + \cdot\cdot + e_nb_n) = v_1b_1 + \cdot\cdot + v_nb_n$. Thus the image of f is generated by the columns of A. (See bottom of page 89.)

2) $\{v_1, .., v_n\}$ generates R^m iff f is surjective iff (for any $C \in R^m$, $AX = C$ has a solution).

3) $\{v_1, .., v_n\}$ is independent iff f is injective iff $AX = \underline{0}$ has a unique solution iff ($\exists\ C \in R^m$ such that $AX = C$ has a unique solution).

4) $\{v_1, .., v_n\}$ is a basis for R^m iff f is an isomorphism iff (for any $C \in R^m$, $AX = C$ has a unique solution).

Relating these concepts to square matrices

We now look at the preceding theorem in the special case where $n = m$ and R is a commutative ring. So far in this chapter we have just been cataloging. Now we prove something more substantial, namely that if $f : R^n \to R^n$ is surjective, then f is injective. Later on we will prove that if R is a field, injective implies surjective.

Theorem Suppose R is a commutative ring, $A \in R_n$, and $f : R^n \to R^n$ is defined by $f(B) = AB$. Let $v_1, .., v_n \in R^n$ be the columns of A, and $w_1, .., w_n \in R^n = R_{1,n}$ be the rows of A. Then the following are equivalent.

1) f is an automorphism.

2) A is invertible, i.e., $\mid A \mid$ is a unit in R.

3) $\{v_1, .., v_n\}$ is a basis for R^n.

4) $\{v_1, .., v_n\}$ generates R^n.

5) f is surjective.

2^t) A^t is invertible, i.e., $\mid A^t \mid$ is a unit in R.

3^t) $\{w_1, .., w_n\}$ is a basis for R^n.

4^t) $\{w_1, .., w_n\}$ generates R^n.

Proof Suppose 5) is true and show 2). Since f is onto, $\exists\ u_1, ..., u_n \in R^n$ with $f(u_i) = e_i$. Let $g : R^n \to R^n$ be the homomorphism satisfying $g(e_i) = u_i$. Then $f \circ g$ is the identity. Now g comes from some matrix D and thus $AD = I$. This shows that A has a right inverse and is thus invertible. Recall that the proof of this fact uses determinant, which requires that R be commutative (see the exercise on page 64).

We already know the first three properties are equivalent, 4) and 5) are equivalent, and 3) implies 4). Thus the first five are equivalent. Furthermore, applying this result to A^t shows that the last three properties are equivalent to each other. Since $\mid A \mid = \mid A^t \mid$, 2) and 2^t) are equivalent.

Uniqueness of Dimension

There exists a ring R with $R^2 \approx R^3$ as R-modules, but this is of little interest. If R is commutative, this is impossible, as shown below. First we make a convention.

Convention *For the remainder of this chapter, R will be a commutative ring.*

Theorem If $f : R^m \to R^n$ is a surjective R-module homomorphism, then $m \geq n$.

Proof Suppose $k = n - m$ is positive. Define $h : (R^m \oplus R^k = R^n) \to R^n$ by $h(u, v) = f(u)$. Then h is a surjective homomorphism, and by the previous section, also injective. This is a contradiction and thus $m \geq n$.

Corollary If $f : R^m \to R^n$ is an isomorphism, then $m = n$.

Proof Each of f and f^{-1} is surjective, so $m = n$ by the previous theorem.

Corollary If $\{v_1, .., v_m\}$ generates R^n, then $m \geq n$.

Proof The hypothesis implies there is a surjective homomorphism $R^m \to R^n$. So this follows from the first theorem.

Lemma Suppose M is a f.g. module (i.e., a finite generated R-module). Then if M has a basis, that basis is finite.

Proof Suppose $U \subset M$ is a finite generating set and S is a basis. Then any element of U is a finite linear combination of elements of S, and thus S is finite.

Theorem Suppose M is a f.g. module. If M has a basis, that basis is finite and any other basis has the same number of elements. This number is denoted by $\dim(M)$, the *dimension* of M. (By convention, $\underline{0}$ is a free module of dimension 0.)

Proof By the previous lemma, any basis for M must be finite. M has a basis of n elements iff $M \approx R^n$. The result follows because $R^n \approx R^m$ iff $n = m$.

Change of Basis

Before changing basis, we recall what a basis is. Previously we defined generating, independence, and basis for sequences, not for collections. For the concept of generating it matters not whether you use sequences or collections, but for independence and basis, you must use sequences. Consider the columns of the real matrix $A = \begin{pmatrix} 2 & 3 & 2 \\ 1 & 4 & 1 \end{pmatrix}$. If we consider the column vectors of A as a collection, there are only two of them, yet we certainly don't wish to say the columns of A form a basis for $\mathbf{R}^2$. In a set or collection, there is no concept of repetition. In order to make sense, we must consider the columns of A as an ordered triple of vectors, and this sequence is dependent. In the definition of basis on page 78, basis is defined for sequences, not for sets or collections.

Two sequences cannot begin to be equal unless they have the same index set. Here we follow the classical convention that an index set with n elements will be $\{1, 2, .., n\}$, and thus a basis for M with n elements is a sequence $S = \{u_1, .., u_n\}$ or if you wish, $S = (u_1, .., u_n) \in M^n$. Suppose M is an R-module with a basis of n elements. Recall there is a bijection $\alpha : \text{Hom}_R(R^n, M) \to M^n$ defined by $\alpha(h) = (h(e_1), .., h(e_n))$. Now $h : R^n \to M$ is an isomorphism iff $\alpha(h)$ is a basis for M.

Summary The point of all this is that selecting a basis of n elements for M is the same as selecting an isomorphism from R^n to M, and from this viewpoint, change of basis can be displayed by the diagram below.

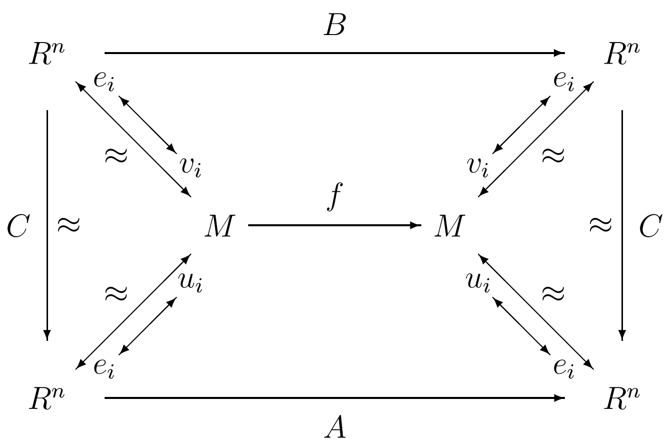
Endomorphisms on R^n are represented by square matrices, and thus have a determinant and trace. Now suppose M is a f.g. free module and $f : M \to M$ is a homomorphism. In order to represent f by a matrix, we must select a basis for M (i.e., an isomorphism with R^n). We will show that this matrix is well defined up to similarity, and thus the determinant, trace, and characteristic polynomial of f are well-defined.

Definition Suppose M is a free module, $S = \{u_1, .., u_n\}$ is a basis for M, and $f : M \to M$ is a homomorphism. The matrix $A = (a_{i,j}) \in R_n$ of f w.r.t. the basis S is defined by $f(u_i) = u_1 a_{1,i} + \cdot\cdot + u_n a_{n,i}$. (Note that if $M = R^n$ and $u_i = e_i$, A is the usual matrix associated with f).

Theorem Suppose $T = \{v_1, .., v_n\}$ is another basis for M and $B \in R_n$ is the matrix of f w.r.t. T. Define $C = (c_{i,j}) \in R_n$ by $v_i = u_1 c_{1,i} + \cdot\cdot + u_n c_{n,i}$. Then C is invertible and $B = C^{-1}AC$, i.e., A and B are similar. Therefore $|A| = |B|$, trace(A)=trace(B), and A and B have the same characteristic polynomial (see page 66 of chapter 4).

Conversely, suppose $C = (c_{i,j}) \in R_n$ is invertible. Define $T = \{v_1, .., v_n\}$ by $v_i = u_1 c_{1,i} + \cdot\cdot + u_n c_{n,i}$. Then T is a basis for M and the matrix of f w.r.t. T is $B = C^{-1}AC$. In other words, conjugation of matrices corresponds to change of basis.

Proof The proof follows by seeing that the following diagram is commutative.



The diagram also explains what it means for A to be the matrix of f w.r.t. the basis S. Let $h : R^n \to M$ be the isomorphism with $h(e_i) = u_i$ for $1 \leq i \leq n$. Then the matrix $A \in R_n$ is the one determined by the endomorphism $h^{-1} \circ f \circ h : R^n \to R^n$. In other words, column i of A is $h^{-1}(f(h(e_i)))$.

An important special case is where $M = R^n$ and $f : R^n \to R^n$ is given by some matrix W. Then h is given by the matrix U whose i^{th} column is u_i and $A = U^{-1}WU$. In other words, W represents f w.r.t. the standard basis, and $U^{-1}WU$ represents f w.r.t. the basis $\{u_1, .., u_n\}$.

Definition Suppose M is a f.g. free module and $f : M \to M$ is a homomorphism. Define $|f|$ to be $|A|$, trace(f) to be trace(A), and $CP_f(x)$ to be $CP_A(x)$, where A is

the matrix of f w.r.t. some basis. By the previous theorem, all three are well-defined, i.e., do not depend upon the choice of basis.

Exercise Let $R = \mathbf{Z}$ and $f : \mathbf{Z}^2 \to \mathbf{Z}^2$ be defined by $f(D) = \begin{pmatrix} 3 & 3 \\ 0 & -1 \end{pmatrix} D$. Find the matrix of f w.r.t. the basis $\left\{ \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 3 \\ 1 \end{pmatrix} \right\}$.

Exercise Let $L \subset \mathbf{R}^2$ be the line $L = \{(r, 2r)^t : r \in \mathbf{R}\}$. Show there is one and only one homomorphism $f : \mathbf{R}^2 \to \mathbf{R}^2$ which is the identity on L and has $f((-1,1)^t) = (1,-1)^t$. Find the matrix $A \in \mathbf{R}_2$ which represents f with respect to the basis $\{(1,2)^t, (-1,1)^t\}$. Find the determinant, trace, and characteristic polynomial of f. Also find the matrix $B \in \mathbf{R}_2$ which represents f with respect to the standard basis. Finally, find an invertible matrix $C \in \mathbf{R}_2$ with $B = C^{-1}AC$.

Vector Spaces

So far in this chapter we have been developing the theory of linear algebra in general. The previous theorem, for example, holds for any commutative ring R, but it must be assumed that the module M is free. Endomorphisms in general will not have a determinant, trace, or characteristic polynomial. We now focus on the case where R is a field F, and show that in this case, every F-module is free. Thus any finitely generated F-module will have a well-defined dimension, and endomorphisms on it will have well-defined determinant, trace, and characteristic polynomial.

In this section, F is a field. F-modules may also be called *vector spaces* and F-module homomorphisms may also be called *linear transformations.*

Theorem Suppose M is an F-module and $v \in M$. Then $v \neq \underline{0}$ iff v is independent. That is, if $v \in V$ and $r \in F$, $vr = \underline{0}$ implies $v = \underline{0}$ in M or $r = \underline{0}$ in F.

Proof Suppose $vr = \underline{0}$ and $r \neq \underline{0}$. Then $\underline{0} = (vr)r^{-1} = v\underline{1} = v$.

Theorem Suppose $M \neq \underline{0}$ is an F-module and $v \in M$. Then v generates M iff v is a basis for M. Furthermore, if these conditions hold, then $M \approx F_F$, any non-zero element of M is a basis, and any two elements of M are dependent.

Proof Suppose v generates M. Then $v \neq \underline{0}$ and is thus independent by the previous theorem. In this case $M \approx F$, and any non-zero element of F is a basis, and any two elements of F are dependent.

Theorem Suppose $M \neq \underline{0}$ is a finitely generated F-module. If $S = \{v_1, .., v_m\}$ generates M, then any maximal independent subsequence of S is a basis for M. Thus any finite independent sequence can be extended to a basis. In particular, M has a finite free basis, and thus is a free F-module.

Proof Suppose, for notational convenience, that $\{v_1, .., v_n\}$ is a maximal independent subsequence of S, and $n < i \leq m$. It must be shown that v_i is a linear combination of $\{v_1, .., v_n\}$. Since $\{v_1, .., v_n, v_i\}$ is dependent, $\exists\ r_1, ..., r_n, r_i$ not all zero, such that $v_1r_1 + \cdots + v_nr_n + v_ir_i = \underline{0}$. Then $r_i \neq \underline{0}$ and $v_i = -(v_1r_1 + \cdots + v_nr_n)r_i^{-1}$. Thus $\{v_1, .., v_n\}$ generates S and thus all of M. Now suppose T is a finite independent sequence. T may be extended to a finite generating sequence, and inside that sequence it may be extended to a maximal independent sequence. Thus T extends to a basis.

After so many routine theorems, it is nice to have one with real power. It not only says any finite independent sequence can be extended to a basis, but it can be extended to a basis inside any finite generating set containing it. This is one of the theorems that makes linear algebra tick. The key hypothesis here is that the ring is a field. If $R = \mathbf{Z}$, then $\mathbf{Z}$ is a free module over itself, and the element 2 of $\mathbf{Z}$ is independent. However it certainly cannot be extended to a basis. Also the finiteness hypothesis in this theorem is only for convenience, as will be seen momentarily.

Since F is a commutative ring, any two bases of M must have the same number of elements, and thus the dimension of M is well defined (see theorem on page 83).

Theorem Suppose M is an F-module of dimension n, and $\{v_1, ..., v_m\}$ is an independent sequence in M. Then $m \leq n$ and if $m = n$, $\{v_1, .., v_m\}$ is a basis.

Proof $\{v_1, .., v_m\}$ extends to a basis with n elements.

The next theorem is just a collection of observations.

Theorem Suppose M and N are finitely generated F-modules.

1) $M \approx F^n$ iff $\dim(M) = n$.
2) $M \approx N$ iff $\dim(M) = \dim(N)$.
3) $F^m \approx F^n$ iff $n = m$.
4) $\dim(M \oplus N) = \dim(M) + \dim(N)$.

Here is the basic theorem for vector spaces in full generality.

Theorem Suppose $M \neq \underline{0}$ is an F-module and $S = \{v_t\}_{t \in T}$ generates M.

1) Any maximal independent subsequence of S is a basis for M.
2) Any independent subsequence of S may be extended to a maximal independent subsequence of S, and thus to a basis for M.
3) Any independent subsequence of M can be extended to a basis for M. In particular, M has a free basis, and thus is a free F-module.

Proof The proof of 1) is the same as in the case where S is finite. Part 2) will follow from the Hausdorff Maximality Principle. An independent subsequence of S is contained in a maximal monotonic tower of independent subsequences. The union of these independent subsequences is still independent, and so the result follows. Part 3) follows from 2) because an independent sequence can always be extended to a generating sequence.

Theorem Suppose M is an F-module and $K \subset M$ is a submodule.

1) K is a summand of M, i.e., $\exists$ a submodule L of M with $K \oplus L = M$.
2) If M is f.g., then $\dim(K) \leq \dim(M)$ and $K = M$ iff $\dim(K) = \dim(M)$.

Proof Let T be a basis for K. Extend T to a basis S for M. Then $S - T$ generates a submodule L with $K \oplus L = M$. Part 2) follows from 1).

Corollary $\mathbf{Q}$ is a summand of $\mathbf{R}$. In other words, $\exists$ a $\mathbf{Q}$-submodule $V \subset \mathbf{R}$ with $\mathbf{Q} \oplus V = \mathbf{R}$ as $\mathbf{Q}$-modules. (See exercise on page 77.)

Proof $\mathbf{Q}$ is a field, $\mathbf{R}$ is a $\mathbf{Q}$-module, and $\mathbf{Q}$ is a submodule of $\mathbf{R}$.

Corollary Suppose M is a f.g. F-module, N is an F-module, and $f : M \to N$ is a homomorphism. Then $\dim(M) = \dim(\ker(f)) + \dim(\text{image}(f))$.

Proof Let $K = \ker(f)$ and $L \subset M$ be a submodule with $K \oplus L = M$. Then $f \mid L : L \to \text{image}(f)$ is an isomorphism.

Exercise Suppose R is a domain with the property that, for R-modules, every submodule is a summand. Show R is a field.

Exercise Find a free $\mathbf{Z}$-module which has a generating set containing no basis.

Exercise The real vector space $\mathbf{R}^2$ is generated by the sequence $S = \{(\pi, 0), (2, 1), (3, 2)\}$. Show there are three maximal independent subsequences of S, and each is a basis for $\mathbf{R}^2$. (Row vectors are used here just for convenience.)

The real vector space $\mathbf{R}^3$ is generated by $S = \{(1, 1, 2), (1, 2, 1), (3, 4, 5), (1, 2, 0)\}$. Show there are three maximal independent subsequences of S and each is a basis for $\mathbf{R}^3$. You may use determinant.

Square matrices over fields

This theorem is just a summary of what we have for square matrices over fields.

Theorem Suppose $A \in F_n$ and $f : F^n \to F^n$ is defined by $f(B) = AB$. Let $v_1, .., v_n \in F^n$ be the columns of A, and $w_1, .., w_n \in F^n = F_{1,n}$ be the rows of A. Then the following are equivalent.

1) $\{v_1, .., v_n\}$ is independent, i.e., f is injective.

2) $\{v_1, .., v_n\}$ is a basis for F^n, i.e., f is an automorphism, i.e., A is invertible, i.e., $\mid A \mid \neq \underline{0}$.

3) $\{v_1, .., v_n\}$ generates F^n, i.e., f is surjective.

1^t) $\{w_1, .., w_n\}$ is independent.

2^t) $\{w_1, .., w_n\}$ is a basis for F^n, i.e., A^t is invertible, i.e., $\mid A^t \mid \neq \underline{0}$.

3^t) $\{w_1, .., w_n\}$ generates F^n.

Proof Except for 1) and 1^t), this theorem holds for any commutative ring R. (See the section **Relating these concepts to square matrices**, pages 81 and 82.) Parts 1) and 1^t) follow from the preceding section.

Exercise Add to this theorem more equivalent statements in terms of solutions of n equations in n unknowns.

Overview Suppose each of X and Y is a set with n elements and $f : X \to Y$ is a function. By the pigeonhole principle, f is injective iff f is bijective iff f is surjective. Now suppose each of U and V is a vector space of dimension n and $f : U \to V$ is a linear transformation. It follows from the work done so far that f is injective iff f is bijective iff f is surjective. This shows some of the simple and definitive nature of linear algebra.

Exercise Let $A = (A_1, .., A_n)$ be an $n \times n$ matrix over $\mathbf{Z}$ with column $i = A_i \in \mathbf{Z}^n$. Let $f : \mathbf{Z}^n \to \mathbf{Z}^n$ be defined by $f(B) = AB$ and $\bar{f} : \mathbf{R}^n \to \mathbf{R}^n$ be defined by $\bar{f}(C) = AC$. Show the following are equivalent. (See the exercise on page 79.)

1) $f : \mathbf{Z}^n \to \mathbf{Z}^n$ is injective.

2) The sequence $(A_1, .., A_n)$ is linearly independent over $\mathbf{Z}$.

3) $|A| \neq 0$.

4) $\bar{f} : \mathbf{R}^n \to \mathbf{R}^n$ is injective.

5) The sequence $(A_1, .., A_n)$ is linearly independent over $\mathbf{R}$.

Rank of a matrix Suppose $A \in F_{m,n}$. The row (column) rank of A is defined to be the dimension of the submodule of F^n (F^m) generated by the rows (columns) of A.

Theorem If $C \in F_m$ and $D \in F_n$ are invertible, then the row (column) rank of A is the same as the row (column) rank of CAD.

Proof Suppose $f : F^n \to F^m$ is defined by $f(B) = AB$. Each column of A is a vector in the range F^m, and we know from page 81 that each $f(B)$ is a linear

combination of those vectors. Thus the image of f is the submodule of F^m generated by the columns of A, and its dimension is the column rank of A. This dimension is the same as the dimension of the image of $g \circ f \circ h : F^n \to F^m$, where h is any automorphism on F^n and g is any automorphism on F^m. This proves the theorem for column rank. The theorem for row rank follows using transpose.

Theorem If $A \in F_{m,n}$, the row rank and the column rank of A are equal. This number is called the *rank* of A and is $\leq \min\{m, n\}$.

Proof By the theorem above, elementary row and column operations change neither the row rank nor the column rank. By row and column operations, A may be changed to a matrix H where $h_{1,1} = \cdot\cdot = h_{t,t} = \underline{1}$ and all other entries are $\underline{0}$ (see the first exercise on page 59). Thus row rank $= t =$ column rank.

Exercise Suppose A has rank t. Show that it is possible to select t rows and t columns of A such that the determined $t \times t$ matrix is invertible. Show that the rank of A is the largest integer t such that this is possible.

Exercise Suppose $A \in F_{m,n}$ has rank t. What is the dimension of the solution set of $AX = \underline{0}$?

Definition If N and M are finite dimensional vector spaces and $f : N \to M$ is a linear transformation, the *rank* of f is the dimension of the image of f. If $f : F^n \to F^m$ is given by a matrix A, then the rank of f is the same as the rank of the matrix A.

Geometric Interpretation of Determinant

Suppose $V \subset \mathbf{R}^n$ is some nice subset. For example, if $n = 2$, V might be the interior of a square or circle. There is a concept of the n-dimensional volume of V. For $n = 1$, it is length. For $n = 2$, it is area, and for $n = 3$ it is "ordinary volume". Suppose $A \in \mathbf{R}_n$ and $f : \mathbf{R}^n \to \mathbf{R}^n$ is the homomorphism given by A. The volume of V does not change under translation, i.e., V and $V + p$ have the same volume. Thus $f(V)$ and $f(V+p) = f(V)+f(p)$ have the same volume. In street language, the next theorem says that "f multiplies volume by the absolute value of its determinant".

Theorem The n-dimensional volume of $f(V)$ is $\pm|A|$(the n-dimensional volume of V). Thus if $|A| = \pm 1$, f preserves volume.

Proof If $|A| = 0$, image(f) has dimension $< n$ and thus $f(V)$ has n-dimensional volume 0. If $|A| \neq 0$ then A is the product of elementary matrices (see page 59) and for elementary matrices, the theorem is obvious. The result follows because the determinant of the composition is the product of the determinants.

Corollary If P is the n-dimensional parallelepiped determined by the columns $v_1, .., v_n$ of A, then the n-dimensional volume of P is $\pm|A|$.

Proof Let $V = [0,1] \times \cdot\cdot \times [0,1] = \{e_1t_1 + \cdot\cdot + e_nt_n : 0 \leq t_i \leq 1\}$. Then $P = f(V) = \{v_1t_1 + \cdot\cdot + v_nt_n : 0 \leq t_i \leq 1\}$.

Linear functions approximate differentiable functions locally

We continue with the special case $F = \mathbf{R}$. Linear functions arise naturally in business, science, and mathematics. However this is not the only reason that linear algebra is so useful. It is a central fact that smooth phenomena may be approximated locally by linear phenomena. Without this great simplification, the world of technology as we know it today would not exist. Of course, linear transformations send the origin to the origin, so they must be adjusted by a translation. As a simple example, suppose $h : \mathbf{R} \to \mathbf{R}$ is differentiable and p is a real number. Let $f : \mathbf{R} \to \mathbf{R}$ be the linear transformation $f(x) = h'(p)x$. Then h is approximated near p by $g(x) = h(p) + f(x - p) = h(p) + h'(p)(x - p)$.

Now suppose $V \subset \mathrm{R}^2$ is some nice subset and $h = (h_1, h_2) : V \to \mathbf{R}^2$ is injective and differentiable. Define the Jacobian by $J(h)(x, y) = \begin{pmatrix} \frac{\partial h_1}{\partial x} & \frac{\partial h_1}{\partial y} \\ \frac{\partial h_2}{\partial x} & \frac{\partial h_2}{\partial y} \end{pmatrix}$ and for each $(x, y) \in V$, let $f(x, y) : \mathbf{R}^2 \to \mathbf{R}^2$ be the homomorphism defined by $J(h)(x, y)$. Then for any $(p_1, p_2) \in V$, h is approximated near (p_1, p_2) (after translation) by $f(p_1, p_2)$. The area of V is $\int\int_V 1 dxdy$. From the previous section we know that any homomorphism f multiplies area by $\mid f \mid$. The student may now understand the following theorem from calculus. (Note that if h is the restriction of a linear transformation from $\mathbf{R}^2$ to $\mathbf{R}^2$, this theorem is immediate from the previous section.)

Theorem Suppose the determinant of $J(h)(x, y)$ is non-negative for each $(x, y) \in V$. Then the area of $h(V)$ is $\int\int_V \mid J(h) \mid dxdy$.

The Transpose Principle

We now return to the case where F is a field (of arbitrary characteristic). F-modules may also be called *vector spaces* and submodules may be called *subspaces*. The study of R-modules in general is important and complex. However the study of F-modules is short and simple – every vector space is free and every subspace is a summand. The core of classical linear algebra is not the study of vector spaces, but the study of homomorphisms, and in particular, of endomorphisms. One goal is to show that if $f : V \to V$ is a homomorphism with some given property, there exists a basis of V so that the matrix representing f displays that property in a prominent manner. The next theorem is an illustration of this.

Theorem Let F be a field and n be a positive integer.

1) Suppose V is an n-dimensional vector space and $f : V \to V$ is a homomorphism with $|f| = \underline{0}$. Then $\exists$ a basis of V such that the matrix representing f has its first row zero.
2) Suppose $A \in F_n$ has $|A| = \underline{0}$. Then $\exists$ an invertible matrix C such that $C^{-1}AC$ has its first row zero.
3) Suppose V is an n-dimensional vector space and $f : V \to V$ is a homomorphism with $|f| = 0$. Then $\exists$ a basis of V such that the matrix representing f has its first column zero.
4) Suppose $A \in F_n$ has $|A| = \underline{0}$. Then $\exists$ an invertible matrix D such that $D^{-1}AD$ has its first column zero.

We first wish to show that these 4 statements are equivalent. We know that 1) and 2) are equivalent and also that 3) and 4) are equivalent because change of basis corresponds to conjugation of the matrix. Now suppose 2) is true and show 4) is true. Suppose $|A| = \underline{0}$. Then $|A^t| = \underline{0}$ and by 2) $\exists$ C such that $C^{-1}A^tC$ has first row zero. Thus $(C^{-1}A^tC)^t = C^tA(C^t)^{-1}$ has first row column zero. The result follows by defining $D = (C^t)^{-1}$. Also 4) implies 2).

This is an example of the *transpose principle*. Loosely stated, it is that theorems about change of basis correspond to theorems about conjugation of matrices and theorems about the rows of a matrix correspond to theorems about the columns of a matrix, using transpose. In the remainder of this chapter, this will be used without further comment.

Proof of the theorem We are free to select any of the 4 parts, and we select part 3). Since $\mid f \mid = 0$, f is not injective and $\exists$ a non-zero $v_1 \in V$ with $f(v_1) = \underline{0}$. Now v_1 is independent and extends to a basis $\{v_1, .., v_n\}$. Then the matrix of f w.r.t this basis has first column zero.

Exercise Let $A = \begin{pmatrix} 3\pi & 6 \\ 2\pi & 4 \end{pmatrix}$. Find an invertible matrix $C \in \mathbf{R}_2$ so that $C^{-1}AC$ has first row zero. Also let $A = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 3 & 4 \\ 2 & 1 & 4 \end{pmatrix}$ and find an invertible matrix $D \in \mathbf{R}_3$ so that $D^{-1}AD$ has first column zero.

Exercise Suppose M is an n-dimensional vector space over a field F, k is an integer with $0 < k < n$, and $f : M \to M$ is an endomorphism of rank k. Show there is a basis for M so that the matrix representing f has its first $n-k$ rows zero. Also show there is a basis for M so that the matrix representing f has its first $n-k$ columns zero. Work these out directly without using the transpose principle.

Nilpotent Homomorphisms

In this section it is shown that an endomorphism f is nilpotent iff all of its characteristic roots are $\underline{0}$ iff it may be represented by a strictly upper triangular matrix.

Definition An endomorphism $f : V \to V$ is nilpotent if $\exists$ m with $f^m = \underline{0}$. Any f represented by a strictly upper triangular matrix is nilpotent (see page 56).

Theorem Suppose V is an n-dimensional vector space and $f : V \to V$ is a nilpotent homomorphism. Then $f^n = \underline{0}$ and $\exists$ a basis of V such that the matrix representing f w.r.t. this basis is strictly upper triangular. Thus the characteristic polynomial of f is $CP_f(x) = x^n$.

Proof Suppose $f \neq \underline{0}$ is nilpotent. Let t be the largest positive integer with $f^t \neq \underline{0}$. Then $f^t(V) \subset f^{t-1}(V) \subset \cdot\cdot \subset f(V) \subset V$. Since f is nilpotent, all of these inclusions are proper. Therefore $t < n$ and $f^n = \underline{0}$. Construct a basis for V by starting with a basis for $f^t(V)$, extending it to a basis for $f^{t-1}(V)$, etc. Then the matrix of f w.r.t. this basis is strictly upper triangular.

Note To obtain a matrix which is strictly lower triangular, reverse the order of the basis.

Exercise Use the transpose principle to write 3 other versions of this theorem.

Theorem Suppose V is an n-dimensional vector space and $f : V \to V$ is a homomorphism. Then f is nilpotent iff $CP_f(x) = x^n$. (See the exercise at the end of Chapter 4 for the case $n = 2$.)

Proof Suppose $CP_f(x) = x^n$. For $n = 1$ this implies $f = \underline{0}$, so suppose $n > 1$. Since the constant term of $CP_f(x)$ is $\underline{0}$, the determinant of f is $\underline{0}$. Thus $\exists$ a basis of V such that the matrix A representing f has its first column zero. Let $B \in F_{n-1}$ be the matrix obtained from A by removing its first row and first column. Now $CP_A(x) = x^n = xCP_B(x)$. Thus $CP_B(x) = x^{n-1}$ and by induction on n, B is nilpotent and so $\exists$ C such that $C^{-1}BC$ is strictly upper triangular. Then

$$\begin{pmatrix} 1 & 0 \cdot \cdot 0 \\ 0 & \\ \cdot & C^{-1} \\ \cdot & \\ 0 & \end{pmatrix} \begin{pmatrix} 0 & * \cdot \cdot * \\ \cdot & \\ & B \\ \cdot & \\ 0 & \end{pmatrix} \begin{pmatrix} 1 & 0 \cdot \cdot 0 \\ 0 & \\ \cdot & C \\ \cdot & \\ 0 & \end{pmatrix} = \begin{pmatrix} 0 & * \cdot \cdot * \\ 0 & \\ \cdot & C^{-1}BC \\ \cdot & \\ 0 & \end{pmatrix}$$

is strictly upper triangular.

Exercise Suppose F is a field, $A \in F_3$ is a strictly lower triangular matrix of rank 2, and $B = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$. Using conjugation by elementary matrices, show there is an invertible matrix C so that $C^{-1}AC = B$. Now suppose V is a 3-dimensional vector space and $f : V \to V$ is a nilpotent endomorphism of rank 2. We know f can be represented by a strictly lower triangular matrix. Show there is a basis $\{v_1, v_2, v_3\}$ for V so that B is the matrix representing f. Also show that $f(v_1) = v_2$, $f(v_2) = v_3$, and $f(v_3) = \underline{0}$. In other words, there is a basis for V of the form $\{v, f(v), f^2(v)\}$ with $f^3(v) = \underline{0}$.

Exercise Suppose V is a 3-dimensional vector space and $f : V \to V$ is a nilpotent endomorphism of rank 1. Show there is a basis for V so that the matrix representing f is $\begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$.

Eigenvalues

Our standing hypothesis is that V is an n-dimensional vector space over a field F and $f : V \to V$ is a homomorphism.

Definition An element $\lambda \in F$ is an *eigenvalue* of f if $\exists$ a non-zero $v \in V$ with $f(v) = \lambda v$. Any such v is called an *eigenvector*. $E_\lambda \subset V$ is defined to be the set of all eigenvectors for λ (plus $\underline{0}$). Note that $E_\lambda = \ker(\lambda I - f)$ is a subspace of V. The next theorem shows the eigenvalues of f are just the characteristic roots of f.

Theorem If $\lambda \in F$ then the following are equivalent.

1) λ is an eigenvalue of f, i.e., $(\lambda I - f) : V \to V$ is not injective.
2) $\mid (\lambda I - f) \mid = \underline{0}$.
3) λ is a characteristic root of f, i.e., a root of the characteristic polynomial $CP_f(x) = \mid (xI - A) \mid$, where A is any matrix representing f.

Proof It is immediate that 1) and 2) are equivalent, so let's show 2) and 3) are equivalent. The evaluation map $F[x] \to F$ which sends $h(x)$ to $h(\lambda)$ is a ring homomorphism (see theorem on page 47). So evaluating $(xI - A)$ at $x = \lambda$ and taking determinant gives the same result as taking the determinant of $(xI - A)$ and evaluating at $x = \lambda$. Thus 2) and 3) are equivalent.

The nicest thing you can say about a matrix is that it is similar to a diagonal matrix. Here is one case where that happens.

Theorem Suppose $\lambda_1, .., \lambda_k$ are distinct eigenvalues of f, and v_i is an eigenvector of λ_i for $1 \le i \le k$. Then the following hold.

1) $\{v_1, .., v_k\}$ is independent.

2) If $k = n$, i.e., if $CP_f(x) = (x - \lambda_1) \cdots (x - \lambda_n)$, then $\{v_1, .., v_n\}$ is a basis for V. The matrix of f w.r.t. this basis is the diagonal matrix whose (i, i) term is λ_i.

Proof Suppose $\{v_1, .., v_k\}$ is dependent. Suppose t is the smallest positive integer such that $\{v_1, .., v_t\}$ is dependent, and $v_1 r_1 + \cdots + v_t r_t = \underline{0}$ is a non-trivial linear combination. Note that at least two of the coefficients must be non-zero. Now $(f - \lambda_t)(v_1 r_1 + \cdots + v_t r_t) = v_1(\lambda_1 - \lambda_t) r_1 + \cdots + v_{t-1}(\lambda_{t-1} - \lambda_t) r_{t-1} + \underline{0} = \underline{0}$ is a shorter

non-trivial linear combination. This is a contradiction and proves 1). Part 2) follows from 1) because $\dim(V) = n$.

Exercise Let $A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \in \mathbf{R}_2$. Find an invertible $C \in \mathbf{C}_2$ such that $C^{-1}AC$ is diagonal. Show that C cannot be selected in $\mathbf{R}_2$. Find the characteristic polynomial of A.

Exercise Suppose V is a 3-dimensional vector space and $f : V \to V$ is an endomorphism with $CP_f(x) = (x - \lambda)^3$. Show that $(f - \lambda I)$ has characteristic polynomial x^3 and is thus a nilpotent endomorphism. Show there is a basis for V so that the matrix representing f is $\begin{pmatrix} \lambda & 0 & 0 \\ 1 & \lambda & 0 \\ 0 & 1 & \lambda \end{pmatrix}$, $\begin{pmatrix} \lambda & 0 & 0 \\ 1 & \lambda & 0 \\ 0 & 0 & \lambda \end{pmatrix}$ or $\begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{pmatrix}$.

We could continue and finally give an ad hoc proof of the Jordan canonical form, but in this chapter we prefer to press on to inner product spaces. The Jordan form will be developed in Chapter 6 as part of the general theory of finitely generated modules over Euclidean domains. The next section is included only as a convenient reference.

Jordan Canonical Form

This section should be just skimmed or omitted entirely. It is unnecessary for the rest of this chapter, and is not properly part of the flow of the chapter. The basic facts of Jordan form are summarized here simply for reference.

The statement that a square matrix B over a field F is a *Jordan block* means that $\exists\ \lambda \in F$ such that B is a lower triangular matrix of the form

$$B = \begin{pmatrix} \lambda & & & & 0 \\ 1 & \lambda & & & \\ & & \cdot & & \\ & & & \cdot & \\ 0 & & & 1 & \lambda \end{pmatrix}.$$ B gives a homomorphism $g : F^m \to F^m$ with $g(e_m) = \lambda e_m$

and $g(e_i) = e_{i+1} + \lambda e_i$ for $1 \leq i < m$. Note that $CP_B(x) = (x - \lambda)^m$ and so λ is the only eigenvalue of B, and B satisfies its characteristic polynomial, i.e., $CP_B(B) = \underline{0}$.

Definition A matrix $D \in F_n$ is in Jordan form if $\exists$ Jordan blocks $B_1, .., B_t$ such

that $D = \begin{pmatrix} B_1 & & & & & \\ & B_2 & & 0 & & \\ & & \cdot & & & \\ & & & \cdot & & \\ & 0 & & & & \\ & & & & & B_t \end{pmatrix}$. Suppose D is of this form and $B_i \in F_{n_i}$ has

eigenvalue λ_i. Then $n_1 + \cdot\cdot + n_t = n$ and $CP_D(x) = (x - \lambda_1)^{n_1} \cdot\cdot (x - \lambda_t)^{n_t}$. Note that a diagonal matrix is a special case of Jordan form. D is a diagonal matrix iff each $n_i = 1$, i.e., iff each Jordan block is a 1×1 matrix.

Theorem If $A \in F_n$, the following are equivalent.

1) $\exists$ an invertible $C \in F_n$ such that $C^{-1}AC$ is in Jordan form.

2) $\exists\ \lambda_1, .., \lambda_n \in F$ (not necessarily distinct) such that $CP_A(x) = (x - \lambda_1) \cdot\cdot (x - \lambda_n)$. (In this case we say that all the eigenvalues of A belong to F.)

Theorem Jordan form (when it exists) is unique. This means that if A and D are similar matrices in Jordan form, they have the same Jordan blocks, except possibly in different order.

The reader should use the transpose principle to write three other versions of the first theorem. Also note that we know one special case of this theorem, namely that if A has n distinct eigenvalues in F, then A is similar to a diagonal matrix. Later on it will be shown that if A is a symmetric real matrix, then A is similar to a diagonal matrix.

Let's look at the classical case $A \in \mathbf{R}_n$. The complex numbers are algebraically closed. This means that $CP_A(x)$ will factor completely in $\mathbf{C}[x]$, and thus $\exists\ C \in \mathbf{C}_n$ with $C^{-1}AC$ in Jordan form. C may be selected to be in $\mathbf{R}_n$ iff all the eigenvalues of A are real.

Exercise Find all real matrices in Jordan form that have the following characteristic polynomials: $x(x-2)$, $(x-2)^2$, $(x-2)(x-3)(x-4)$, $(x-2)(x-3)^2$, $(x-2)^2(x-3)^2$, $(x-2)(x-3)^3$.

Exercise Suppose $D \in F_n$ is in Jordan form and has characteristic polynomial $a_0 + a_1 x + \cdot\cdot + x^n$. Show $a_0 I + a_1 D + \cdot\cdot + D^n = \underline{0}$, i.e., show $CP_D(D) = \underline{0}$.

Exercise (Cayley-Hamilton Theorem) Suppose E is a field and $A \in E_n$. Assume the theorem that there is a field F containing E such that $CP_A(x)$ factors completely in $F[x]$. Thus $\exists$ an invertible $C \in F_n$ such that $D = C^{-1}AC$ is in Jordan form. Use this to show $CP_A(A) = \underline{0}$. (See the second exercise on page 66.)

Exercise Suppose $A \in F_n$ is in Jordan form. Show A is nilpotent iff $A^n = \underline{0}$ iff $CP_A(x) = x^n$. (Note how easy this is in Jordan form.)

Inner Product Spaces

The two most important fields for mathematics and science in general are the real numbers and the complex numbers. Finitely generated vector spaces over **R** or **C** support inner products and are thus geometric as well as algebraic objects. The theories for the real and complex cases are quite similar, and both could have been treated here. However, for simplicity, attention is restricted to the case $F = \mathbf{R}$. In the remainder of this chapter, the power and elegance of linear algebra become transparent for all to see.

Definition Suppose V is a real vector space. An *inner product* (or *dot product*) on V is a function $V \times V \to \mathbf{R}$ which sends (u, v) to $u \cdot v$ and satisfies

1) $(u_1r_1 + u_2r_2) \cdot v = (u_1 \cdot v)r_1 + (u_2 \cdot v)r_2$ for all $u_1, u_2, v \in V$
$v \cdot (u_1r_1 + u_2r_2) = (v \cdot u_1)r_1 + (v \cdot u_2)r_2$ and $r_1, r_2 \in \mathbf{R}$.

2) $u \cdot v = v \cdot u$ for all $u, v \in V$.

3) $u \cdot u \geq 0$ and $u \cdot u = 0$ iff $u = \underline{0}$ for all $u \in V$.

Theorem Suppose V has an inner product.

1) If $v \in V$, $f : V \to \mathbf{R}$ defined by $f(u) = u \cdot v$ is a homomorphism. Thus $\underline{0} \cdot v = 0$.

2) Schwarz' inequality. If $u, v \in V$, $(u \cdot v)^2 \leq (u \cdot u)(v \cdot v)$.

Proof of 2) Let $a = \sqrt{v \cdot v}$ and $b = \sqrt{u \cdot u}$. If a or b is 0, the result is obvious. Suppose neither a nor b is 0. Now $0 \leq (ua \pm vb) \cdot (ua \pm vb) = (u \cdot u)a^2 \pm 2ab(u \cdot v) + (v \cdot v)b^2 = b^2a^2 \pm 2ab(u \cdot v) + a^2b^2$. Dividing by $2ab$ yields $0 \leq ab \pm (u \cdot v)$ or $| u \cdot v | \leq ab$.

Theorem Suppose V has an inner product. Define the *norm* or *length* of a vector v by $\|v\| = \sqrt{v \cdot v}$. The following properties hold.

1) $\|v\| = 0$ iff $v = \underline{0}$.

2) $\|vr\| = \|v\| \mid r \mid$.

3) $\mid u \cdot v \mid \leq \|u\|\|v\|$. (Schwarz' inequality)

4) $\|u + v\| \leq \|u\| + \|v\|$. (The triangle inequality)

Proof of 4) $\|u + v\|^2 = (u + v) \cdot (u + v) = \|u\|^2 + 2(u \cdot v) + \|v\|^2 \leq \|u\|^2 + 2\|u\|\|v\| + \|v\|^2 = (\|u\| + \|v\|)^2$.

Definition An Inner Product Space (IPS) is a real vector space with an inner product. Suppose V is an IPS. A sequence $\{v_1, .., v_n\}$ is *orthogonal* provided $v_i \cdot v_j = 0$ when $i \neq j$. The sequence is *orthonormal* if it is orthogonal and each vector has length 1, i.e., $v_i \cdot v_j = \delta_{i,j}$ for $1 \leq i, j \leq n$.

Theorem If $S = \{v_1, .., v_n\}$ is an orthogonal sequence of non-zero vectors in an IPS V, then S is independent. Furthermore $\left\{ \dfrac{v_1}{\|v_1\|}, \cdots, \dfrac{v_n}{\|v_n\|} \right\}$ is orthonormal.

Proof Suppose $v_1 r_1 + \cdots + v_n r_n = \underline{0}$. Then $0 = (v_1 r_1 + \cdots + v_n r_n) \cdot v_i = r_i (v_i \cdot v_i)$ and thus $r_i = 0$. Thus S is independent. The second statement is transparent.

It is easy to define an inner product, as is shown by the following theorem.

Theorem Suppose V is a real vector space with a basis $S = \{v_1, .., v_n\}$. Then there is a unique inner product on V which makes S an orthornormal basis. It is given by the formula $(v_1 r_1 + \cdots + v_n r_n) \cdot (v_1 s_1 + \cdots + v_n s_n) = r_1 s_1 + \cdots + r_n s_n$.

Convention $\mathbf{R}^n$ will be assumed to have the *standard inner product* defined by $(r_1, .., r_n)^t \cdot (s_1, .., s_n)^t = r_1 s_1 + \cdots + r_n s_n$. $S = \{e_1, .., e_n\}$ will be called the *canonical* or *standard orthonormal basis* (see page 72). The next theorem shows that this inner product has an amazing geometry.

Theorem If $u, v \in \mathbf{R}^n$, $u \cdot v = \|u\|\|v\| \cos \Theta$ where Θ is the angle between u

and v.

Proof Let $u = (r_1, .., r_n)$ and $v = (s_1, .., s_n)$. By the law of cosines $\|u - v\|^2 = \|u\|^2 + \|v\|^2 - 2\|u\|\|v\| \cos \Theta$. So $(r_1 - s_1)^2 + \cdot\cdot + (r_n - s_n)^2 = r_1^2 + \cdot\cdot + r_n^2 + s_1^2 + \cdot\cdot + s_n^2 - 2\|u\|\|v\| \cos \Theta$. Thus $r_1 s_1 + \cdot\cdot + r_n s_n = \|u\|\|v\| \cos \Theta$.

Exercise This is a simple exercise to observe that hyperplanes in $\mathbf{R}^n$ are cosets. Suppose $f : \mathbf{R}^n \to \mathbf{R}$ is a non-zero homomorphism given by a matrix $A = (a_1, .., a_n) \in \mathbf{R}_{1,n}$. Then $L = \ker(f)$ is the set of all solutions to $a_1 x_1 + \cdot\cdot + a_n x_n = 0$, i.e., the set of all vectors perpendicular to A. Now suppose $b \in \mathbf{R}$ and $C = \begin{pmatrix} c_1 \\ \cdot \\ \cdot \\ c_n \end{pmatrix} \in \mathbf{R}^n$ has $f(C) = b$. Then $f^{-1}(b)$ is the set of all solutions to $a_1 x_1 + \cdot\cdot + a_n x_n = b$ which is the coset $L + C$, and this the set of all solutions to $a_1(x_1 - c_1) + \cdot\cdot + a_n(x_n - c_n) = 0$.

Gram-Schmidt orthonormalization

Theorem (Fourier series) Suppose W is an IPS with an orthonormal basis $\{w_1, .., w_n\}$. Then if $v \in W$, $v = w_1(v \cdot w_1) + \cdot\cdot + w_n(v \cdot w_n)$.

Proof $v = w_1 r_1 + \cdot\cdot + w_n r_n$ and $v \cdot w_i = (w_1 r_1 + \cdot\cdot + w_n r_n) \cdot w_i = r_i$

Theorem Suppose W is an IPS, $Y \subset W$ is a subspace with an orthonormal basis $\{w_1, .., w_k\}$, and $v \in W - Y$. Define the *projection* of v onto Y by $p(v) = w_1(v \cdot w_1) + \cdot\cdot + w_k(v \cdot w_k)$, and let $w = v - p(v)$. Then $(w \cdot w_i) = (v - w_1(v \cdot w_1) \cdot\cdot - w_k(v \cdot w_k)) \cdot w_i = 0$. Thus if $w_{k+1} = \dfrac{w}{\|w\|}$, then $\{w_1, .., w_{k+1}\}$ is an orthonormal basis for the subspace generated by $\{w_1, .., w_k, v\}$. If $\{w_1, .., w_k, v\}$ is already orthonormal, $w_{k+1} = v$.

Theorem **(Gram-Schmidt)** Suppose W is an IPS with a basis $\{v_1, .., v_n\}$. Then W has an orthonormal basis $\{w_1, .., w_n\}$. Moreover, any orthonormal sequence in W extends to an orthonormal basis of W.

Proof Let $w_1 = \dfrac{v_1}{\|v_1\|}$. Suppose inductively that $\{w_1, .., w_k\}$ is an orthonormal basis for Y, the subspace generated by $\{v_1, .., v_k\}$. Let $w = v_{k+1} - p(v_{k+1})$ and

$w_{k+1} = \dfrac{w}{\|w\|}$. Then by the previous theorem, $\{w_1, .., w_{k+1}\}$ is an orthonormal basis for the subspace generated by $\{w_1, .., w_k, v_{k+1}\}$. In this manner an orthonormal basis for W is constructed. Notice that this construction defines a function h which sends a basis for W to an orthonormal basis for W (see topology exercise on page 103).

Now suppose W has dimension n and $\{w_1, .., w_k\}$ is an orthonormal sequence in W. Since this sequence is independent, it extends to a basis $\{w_1, .., w_k, v_{k+1}, .., v_n\}$. The process above may be used to modify this to an orthonormal basis $\{w_1, .., w_n\}$.

Exercise Let $f : \mathbf{R}^3 \to \mathbf{R}$ be the homomorphism defined by the matrix (2,1,3). Find an orthonormal basis for the kernel of f. Find the projection of $(e_1 + e_2)$ onto $\ker(f)$. Find the angle between $e_1 + e_2$ and the plane $\ker(f)$.

Exercise Let $W = \mathbf{R}^3$ have the standard inner product and $Y \subset W$ be the subspace generated by $\{w_1, w_2\}$ where $w_1 = (1,0,0)^t$ and $w_2 = (0,1,0)^t$. W is generated by the sequence $\{w_1, w_2, v\}$ where $v = (1,2,3)^t$. As in the first theorem of this section, let $w = v - p(v)$, where $p(v)$ is the projection of v onto Y, and set $w_3 = \dfrac{w}{\|w\|}$. Find w_3 and show that for any t with $0 \le t \le 1$, $\{w_1, w_2, (1-t)v + tw_3\}$ is a basis for W. This is a key observation for an exercise on page 103 showing $O(n)$ is a deformation retract of $GL_n(\mathbf{R})$.

Isometries Suppose each of U and V is an IPS. A homomorphism $f : U \to V$ is said to be an *isometry* provided it is an isomorphism and for any u_1, u_2 in U, $(u_1 \cdot u_2)_U = (f(u_1) \cdot f(u_2))_V$.

Theorem Suppose each of U and V is an n-dimensional IPS, $\{u_1, .., u_n\}$ is an orthonormal basis for U, and $f : U \to V$ is a homomorphism. Then f is an isometry iff $\{f(u_1), .., f(u_n)\}$ is an orthonormal sequence in V.

Proof Isometries certainly preserve orthonormal sequences. So suppose $T = \{f(u_1), .., f(u_n)\}$ is an orthonormal sequence in V. Then T is independent and thus T is a basis for V and thus f is an isomorphism (see the second theorem on page 79). It is easy to check that f preserves inner products.

We now come to one of the definitive theorems in linear algebra. It is that, up to isometry, there is only one inner product space for each dimension.

Theorem Suppose each of U and V is an n-dimensional IPS. Then $\exists$ an isometry $f: U \to V$. In particular, U is isometric to $\mathbf{R}^n$ with its standard inner product.

Proof There exist orthonormal bases $\{u_1, .., u_n\}$ for U and $\{v_1, .., v_n\}$ for V. By the first theorem on page 79, there exists a homomorphism $f : U \to V$ with $f(u_i) = v_i$, and by the previous theorem, f is an isometry.

Exercise Let $f : \mathbf{R}^3 \to \mathbf{R}$ be the homomorphism defined by the matrix (2,1,3). Find a linear transformation $h : \mathbf{R}^2 \to \mathbf{R}^3$ which gives an isometry from $\mathbf{R}^2$ to $\ker(f)$.

Orthogonal Matrices

As noted earlier, linear algebra is not so much the study of vector spaces as it is the study of endomorphisms. We now wish to study isometries from $\mathbf{R}^n$ to $\mathbf{R}^n$.

We know from a theorem on page 90 that an endomorphism preserves volume iff its determinant is ± 1. Isometries preserve inner product, and thus preserve angle and distance, and so certainly preserve volume.

Theorem Suppose $A \in \mathbf{R}_n$ and $f : \mathbf{R}^n \to \mathbf{R}^n$ is the homomorphism defined by $f(B) = AB$. Then the following are equivalent.

1) The columns of A form an orthonormal basis for $\mathbf{R}^n$, i.e., $A^t A = I$.
2) The rows of A form an orthonormal basis for $\mathbf{R}^n$, i.e., $AA^t = I$.
3) f is an isometry.

Proof A left inverse of a matrix is also a right inverse (see the exercise on page 64). Thus 1) and 2) are equivalent because each of them says A is invertible with $A^{-1} = A^t$. Now $\{e_1, .., e_n\}$ is the canonical orthonormal basis for $\mathbf{R}^n$, and $f(e_i)$ is column i of A. Thus by the previous section, 1) and 3) are equivalent.

Definition If $A \in \mathbf{R}_n$ satisfies these three conditions, A is said to be *orthogonal*. The set of all such A is denoted by $O(n)$, and is called the *orthogonal group*.

Theorem

1) If A is orthogonal, $\mid A \mid = \pm 1$.

2) If A is orthogonal, A^{-1} is orthogonal. If A and C are orthogonal, AC is orthogonal. Thus $O(n)$ is a multiplicative subgroup of $GL_n(\mathbf{R})$.

3) Suppose A is orthogonal and f is defined by $f(B) = AB$. Then f preserves distances and angles. This means that if $u, v \in \mathbf{R}^n$, $\|u - v\| = \|f(u) - f(v)\|$ and the angle between u and v is equal to the angle between $f(u)$ and $f(v)$.

Proof Part 1) follows from $|A|^2 = |A|\,|A^t| = |I| = 1$. Part 2) is immediate, because isometries clearly form a subgroup of the multiplicative group of all automorphisms. For part 3) assume $f : \mathbf{R}^n \to \mathbf{R}^n$ is an isometry. Then $\|u - v\|^2 = (u - v) \cdot (u - v) = f(u - v) \cdot f(u - v) = \|f(u - v)\|^2 = \|f(u) - f(v)\|^2$. The proof that f preserves angles follows from $u \cdot v = \|u\|\|v\|\cos\Theta$.

Exercise Show that if $A \in O(2)$ has $|A| = 1$, then $A = \begin{pmatrix} \cos\Theta & -\sin\Theta \\ \sin\Theta & \cos\Theta \end{pmatrix}$ for some number Θ. (See the exercise on page 56.)

Exercise (topology) Let $\mathbf{R}_n \approx \mathbf{R}^{n^2}$ have its usual metric topology. This means a sequence of matrices $\{A_i\}$ converges to A iff it converges coordinatewise. Show $GL_n(\mathbf{R})$ is an open subset and $O(n)$ is closed and compact. Let $h : GL_n(\mathbf{R}) \to O(n)$ be defined by Gram-Schmidt. Show $H : GL_n(\mathbf{R}) \times [0,1] \to GL_n(\mathbf{R})$ defined by $H(A,t) = (1-t)A + th(A)$ is a deformation retract of $GL_n(R)$ to $O(n)$.

Diagonalization of Symmetric Matrices

We continue with the case $F = \mathbf{R}$. Our goals are to prove that, if A is a symmetric matrix, all of its eigenvalues are real and that $\exists$ an orthogonal matrix C such that $C^{-1}AC$ is diagonal. As background, we first note that symmetric is the same as self-adjoint.

Theorem Suppose $A \in \mathbf{R}_n$ and $u, v \in \mathbf{R}^n$. Then $(A^t u) \cdot v = u \cdot (Av)$.

Proof If $y, z \in \mathbf{R}^n$, then the dot product $y \cdot z$, is the matrix product $y^t z$, and matrix multiplication is associative. Thus $(A^t u) \cdot v = (u^t A)v = u^t(Av) = u \cdot (Av)$.

Definition Suppose $A \in \mathbf{R}_n$. A is said to be *symmetric* provided $A^t = A$. Note that any diagonal matrix is symmetric. A is said to be *self-adjoint* if $(Au)\cdot v = u\cdot(Av)$ for all $u, v \in \mathbf{R}^n$. The next theorem is just an exercise using the previous theorem.

Theorem A is symmetric iff A is self-adjoint.

Theorem Suppose $A \in \mathbf{R}_n$ is symmetric. Then $\exists$ real numbers $\lambda_1, .., \lambda_n$ (not necessarily distinct) such that $CP_A(x) = (x - \lambda_1)(x - \lambda_2) \cdots (x - \lambda_n)$. That is, all the eigenvalues of A are real.

Proof We know $CP_A(x)$ factors into linears over $\mathbf{C}$. If $\mu = a + bi$ is a complex number, its conjugate is defined by $\bar{\mu} = a - bi$. If $h : \mathbf{C} \to \mathbf{C}$ is defined by $h(\mu) = \bar{\mu}$, then h is a ring isomorphism which is the identity on $\mathbf{R}$. If $w = (a_{i,j})$ is a complex matrix or vector, its conjugate is defined by $\bar{w} = (\bar{a}_{i,j})$. Since $A \in \mathbf{R}_n$ is a real symmetric matrix, $A = A^t = \bar{A}^t$. Now suppose λ is a complex eigenvalue of A and $v \in \mathbf{C}^n$ is an eigenvector with $Av = \lambda v$. Then $\lambda(v^t \bar{v}) = (\lambda v)^t \bar{v} = (Av)^t \bar{v} = (v^t A)\bar{v} = v^t(A\bar{v}) = v^t(\overline{Av}) = v^t(\overline{\lambda v}) = \bar{\lambda}(v^t \bar{v})$. Thus $\lambda = \bar{\lambda}$ and $\lambda \in \mathbf{R}$. Or you can define a complex inner product on $\mathbf{C}^n$ by $(w \cdot v) = w^t \bar{v}$. The proof then reads as $\lambda(v \cdot v) = (\lambda v \cdot v) = (Av \cdot v) = (v \cdot Av) = (v \cdot \lambda v) = \bar{\lambda}(v \cdot v)$. Either way, λ is a real number.

We know that eigenvectors belonging to distinct eigenvalues are linearly independent. For symmetric matrices, we show more, namely that they are perpendicular.

Theorem Suppose A is symmetric, $\lambda_1, \lambda_2 \in \mathbf{R}$ are distinct eigenvalues of A, and $Au = \lambda_1 u$ and $Av = \lambda_2 v$. Then $u \cdot v = 0$.

Proof $\lambda_1(u \cdot v) = (Au) \cdot v = u \cdot (Av) = \lambda_2(u \cdot v)$.

Review Suppose $A \in \mathbf{R}_n$ and $f : \mathbf{R}^n \to \mathbf{R}^n$ is defined by $f(B) = AB$. Then A represents f w.r.t. the canonical orthonormal basis. Let $S = \{v_1, .., v_n\}$ be another basis and $C \in \mathbf{R}_n$ be the matrix with v_i as column i. Then $C^{-1}AC$ is the matrix representing f w.r.t. S. Now S is an orthonormal basis iff C is an orthogonal matrix.

Summary Representing f w.r.t. an orthonormal basis is the same as conjugating A by an orthogonal matrix.

Theorem Suppose $A \in \mathbf{R}_n$ and $C \in O(n)$. Then A is symmetric iff $C^{-1}AC$ is symmetric.

Proof Suppose A is symmetric. Then $(C^{-1}AC)^t = C^t A (C^{-1})^t = C^{-1}AC$.

The next theorem has geometric and physical implications, but for us, just the incredibility of it all will suffice.

Theorem If $A \in \mathbf{R}_n$, the following are equivalent.

1) A is symmetric.
2) $\exists\ C \in O(n)$ such that $C^{-1}AC$ is diagonal.

Proof By the previous theorem, 2) $\Rightarrow$ 1). Show 1) $\Rightarrow$ 2). Suppose A is a symmetric 2×2 matrix. Let λ be an eigenvalue for A and $\{v_1, v_2\}$ be an orthonormal basis for $\mathbf{R}^2$ with $Av_1 = \lambda v_1$. Then w.r.t this basis, the transformation determined by A is represented by $\begin{pmatrix} \lambda & b \\ 0 & d \end{pmatrix}$. Since this matrix is symmetric, $b = 0$.

Now suppose by induction that the theorem is true for symmetric matrices in $\mathbf{R}_t$ for $t < n$, and suppose A is a symmetric $n \times n$ matrix. Denote by $\lambda_1, .., \lambda_k$ the distinct eigenvalues of A, $k \leq n$. If $k = n$, the proof is immediate, because then there is a basis of eigenvectors of length 1, and they must form an orthonormal basis. So suppose $k < n$. Let $v_1, .., v_k$ be eigenvectors for $\lambda_1, .., \lambda_k$ with each $\| v_i \| = 1$. They may be extended to an orthonormal basis $v_1, .., v_n$. With respect to this basis, the transformation determined by A is represented by

$$\begin{pmatrix} \begin{pmatrix} \lambda_1 & & & \\ & \cdot & & \\ & & \cdot & \\ & & & \lambda_k \end{pmatrix} & (B) \\ & \\ (0) & (D) \end{pmatrix}.$$

Since this is a symmetric matrix, $B = 0$ and D is a symmetric matrix of smaller size. By induction, $\exists$ an orthogonal C such that $C^{-1}DC$ is diagonal. Thus conjugating by $\begin{pmatrix} I & 0 \\ 0 & C \end{pmatrix}$ makes the entire matrix diagonal.

This theorem is so basic we state it again in different terminology. If V is an IPS, a linear transformation $f : V \to V$ is said to be self-adjoint provided $(u \cdot f(v)) = (f(u) \cdot v)$ for all $u, v \in V$.

Theorem If V is an n-dimensional IPS and $f : V \to V$ is a linear transformation, then the following are equivalent.

1) f is self-adjoint.
2) $\exists$ an orthonormal basis $\{v_1, ..., v_n\}$ for V with each v_i an eigenvector of f.

Exercise Let $A = \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix}$. Find an orthogonal C such that $C^{-1}AC$ is diagonal. Do the same for $A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$.

Exercise Suppose $A, D \in \mathbf{R}_n$ are symmetric. Under what conditions are A and D similar? Show that, if A and D are similar, $\exists$ an orthogonal C such that $D = C^{-1}AC$.

Exercise Suppose V is an n-dimensional real vector space. We know that V is isomorphic to $\mathbf{R}^n$. Suppose f and g are isomorphisms from V to $\mathbf{R}^n$ and A is a subset of V. Show that $f(A)$ is an open subset of $\mathbf{R}^n$ iff $g(A)$ is an open subset of $\mathbf{R}^n$. This shows that V, an algebraic object, has a god-given topology. Of course, if V has an inner product, it automatically has a metric, and this metric will determine that same topology. Finally, suppose V and W are finite-dimensional real vector spaces and $h : V \to W$ is a linear transformation. Show that h is continuous.

Exercise Define $E : \mathbf{C}_n \to \mathbf{C}_n$ by $E(A) = e^A = I + A + (1/2!)A^2 + \cdots$. This series converges and thus E is a well defined function. If $AB = BA$, then $E(A+B) = E(A)E(B)$. Since A and $-A$ commute, $I = E(\underline{0}) = E(A - A) = E(A)E(-A)$, and thus $E(A)$ is invertible with $E(A)^{-1} = E(-A)$. Furthermore $E(A^t) = E(A)^t$, and if C is invertible, $E(C^{-1}AC) = C^{-1}E(A)C$. Now use the results of this section to prove the statements below. (For part 1, assume the Jordan form, i.e., assume any $A \in \mathbf{C}_n$ is similar to a lower triangular matrix.)

1) If $A \in \mathbf{C}_n$, then $\mid e^A \mid = e^{\text{trace}(A)}$. Thus if $A \in \mathbf{R}_n$, $\mid e^A \mid = 1$ iff trace$(A) = 0$.

2) $\exists$ a non-zero matrix $N \in \mathbf{R}_2$ with $e^N = I$.

3) If $N \in \mathbf{R}_n$ is symmetric, then $e^N = I$ iff $N = \underline{0}$.

4) If $A \in \mathbf{R}_n$ and $A^t = -A$, then $e^A \in O(n)$.

Chapter 6

Appendix

The five previous chapters were designed for a year undergraduate course in algebra. In this appendix, enough material is added to form a basic first year graduate course. Two of the main goals are to characterize finitely generated abelian groups and to prove the Jordan canonical form. The style is the same as before, i.e., everything is right down to the nub. The organization is mostly a linearly ordered sequence except for the last two sections on determinants and dual spaces. These are independent sections added on at the end.

Suppose R is a commutative ring. An R-module M is said to be cyclic if it can be generated by one element, i.e., $M \approx R/I$ where I is an ideal of R. The basic theorem of this chapter is that if R is a Euclidean domain and M is a finitely generated R-module, then M is the sum of cyclic modules. Thus if M is torsion free, it is a free R-module. Since $\mathbf{Z}$ is a Euclidean domain, finitely generated abelian groups are the sums of cyclic groups – one of the jewels of abstract algebra.

Now suppose F is a field and V is a finitely generated F-module. If $T : V \to V$ is a linear transformation, then V becomes an $F[x]$-module by defining $vx = T(v)$. Now $F[x]$ is a Euclidean domain and so $V_{F[x]}$ is the sum of cyclic modules. This classical and very powerful technique allows an easy proof of the canonical forms. There is a basis for V so that the matrix representing T is in Rational canonical form. If the characteristic polynomial of T factors into the product of linear polynomials, then there is a basis for V so that the matrix representing T is in Jordan canonical form. This always holds if $F = \mathbf{C}$. A matrix in Jordan form is a lower triangular matrix with the eigenvalues of T displayed on the diagonal, so this is a powerful concept.

In the chapter on matrices, it is stated without proof that the determinant of the product is the product of the determinants. A proof of this, which depends upon the classification of certain types of alternating multilinear forms, is given in this chapter. The final section gives the fundamentals of dual spaces.

The Chinese Remainder Theorem

On page 50 in the chapter on rings, the Chinese Remainder Theorem was proved for the ring of integers. In this section this classical topic is presented in full generality. Surprisingly, the theorem holds even for non-commutative rings.

Definition Suppose R is a ring and $A_1, A_2, ..., A_m$ are ideals of R. Then the *sum* $A_1 + A_2 + \cdots + A_m$ is the set of all $a_1 + a_2 + \cdots + a_m$ with $a_i \in A_i$. The *product* $A_1 A_2 \cdots A_m$ is the set of all finite sums of elements $a_1 a_2 \cdots a_m$ with $a_i \in A_i$. Note that the sum and product of ideals are ideals and $A_1 A_2 \cdots A_m \subset (A_1 \cap A_2 \cap \cdots \cap A_m)$.

Definition Ideals A and B of R are said to be *comaximal* if $A + B = R$.

Theorem If A and B are ideals of a ring R, then the following are equivalent.

1) A and B are comaximal.
2) $\exists\ a \in A$ and $b \in B$ with $a + b = \underline{1}$.
3) $\pi(A) = R/B$ where $\pi : R \to R/B$ is the projection.

Theorem If $A_1, A_2, ..., A_m$ and B are ideals of R with A_i and B comaximal for each i, then $A_1 A_2 \cdots A_m$ and B are comaximal. Thus $A_1 \cap A_2 \cap \cdots \cap A_m$ and B are comaximal.

Proof Consider $\pi : R \to R/B$. Then $\pi(A_1 A_2 \cdots A_m) = \pi(A_1)\pi(A_2) \cdots \pi(A_m) = (R/B)(R/B) \cdots (R/B) = R/B$.

Chinese Remainder Theorem Suppose $A_1, A_2, ..., A_n$ are pairwise comaximal ideals of R, with each $A_i \neq R$. Then the natural map $\pi : R \to R/A_1 \times R/A_2 \times \cdots \times R/A_n$ is a surjective ring homomorphism with kernel $A_1 \cap A_2 \cap \cdots \cap A_n$.

Proof There exists $a_i \in A_i$ and $b_i \in A_1 A_2 \cdots A_{i-1} A_{i+1} \cdots A_n$ with $a_i + b_i = \underline{1}$. Note that $\pi(b_i) = (0, .., 0, \underline{1}_i, 0, .., 0)$. If $(r_1 + A_1, r_2 + A_2, ..., r_n + A_n)$ is an element of the range, it is the image of $r_1 b_1 + r_2 b_2 + \cdots + r_n b_n = r_1(\underline{1} - a_1) + r_2(\underline{1} - a_2) + \cdots + r_n(\underline{1} - a_n)$.

Theorem If R is commutative and $A_1, A_2, ..., A_n$ are pairwise comaximal ideals of R, then $A_1 A_2 \cdots A_n = A_1 \cap A_2 \cap \cdots \cap A_n$.

Proof for $n = 2$. Show $A_1 \cap A_2 \subset A_1 A_2$. $\exists\ a_1 \in A_1$ and $a_2 \in A_2$ with $a_1 + a_2 = \underline{1}$. If $c \in A_1 \cap A_2$, then $c = c(a_1 + a_2) \in A_1 A_2$.

Prime and Maximal Ideals and UFD^s

In the first chapter on background material, it was shown that **Z** is a unique factorization domain. Here it will be shown that this property holds for any principle ideal domain. Later on it will be shown that every Euclidean domain is a principle ideal domain. Thus every Euclidean domain is a unique factorization domain.

Definition Suppose R is a commutative ring and $I \subset R$ is an ideal.

I is *prime* means $I \neq R$ and if $a, b \in R$ have $ab \in I$, then a or $b \in I$.
I is *maximal* means $I \neq R$ and there are no ideals properly between I and R.

Theorem $\underline{0}$ is a prime ideal of R iff R is ____________
$\underline{0}$ is a maximal ideal of R iff R is ____________

Theorem Suppose $J \subset R$ is an ideal, $J \neq R$.
J is a prime ideal iff R/J is ____________
J is a maximal ideal iff R/J is ____________

Corollary Maximal ideals are prime.

Proof Every field is a domain.

Theorem If $a \in R$ is not a unit, then $\exists$ a maximal ideal I of R with $a \in I$.

Proof This is a classical application of the Hausdorff Maximality Principle. Consider $\{J : J$ is an ideal of R containing a with $J \neq R\}$. This collection contains a maximal monotonic collection $\{V_t\}_{t \in T}$. The ideal $V = \bigcup\limits_{t \in T} V_t$ does not contain $\underline{1}$ and thus is not equal to R. Therefore V is equal to some V_t and is a maximal ideal containing a.

Note To properly appreciate this proof, the student should work the exercise in group theory at the end of this section (see page 114).

Definition Suppose R is a domain and $a, b \in R$. Then we say $a \sim b$ iff there exists a unit u with $au = b$. Note that $\sim$ is an equivalence relation. If $a \sim b$, then a

and b are said to be *associates*.

Examples If R is a domain, the associates of $\underline{1}$ are the units of R, while the only associate of $\underline{0}$ is $\underline{0}$ itself. If $n \in \mathbf{Z}$ is not zero, then its associates are n and $-n$. If F is a field and $g \in F[x]$ is a non-zero polynomial, then the associates of g are all cg where c is a non-zero constant.

The following theorem is elementary, but it shows how associates fit into the scheme of things. An element a divides b $(a|b)$ if $\exists!$ $c \in R$ with $ac = b$.

Theorem Suppose R is a domain and $a, b \in (R - \underline{0})$. Then the following are equivalent.

1) $a \sim b$.
2) $a|b$ and $b|a$.
3) $aR = bR$.

Parts 1) and 3) above show there is a bijection from the associate classes of R to the principal ideals of R. Thus if R is a PID, there is a bijection from the associate classes of R to the ideals of R. If an element of a domain generates a non-zero prime ideal, it is called a prime element.

Definition Suppose R is a domain and $a \in R$ is a non-zero non-unit.

1) a is *irreducible* if it does not factor, i.e., $a = bc \Rightarrow b$ or c is a unit.
2) a is *prime* if it generates a prime ideal, i.e., $a|bc \Rightarrow a|b$ or $a|c$.

Note If a is a prime and $a|c_1c_2 \cdots c_n$, then $a|c_i$ for some i. This follows from the definition and induction on n. If each c_j is irreducible, then $a \sim c_i$ for some i.

Note If $a \sim b$, then a is irreducible (prime) iff b is irreducible (prime). In other words, if a is irreducible (prime) and u is a unit, then au is irreducible (prime).

Note a is prime $\Rightarrow a$ is irreducible. This is immediate from the definitions.

Theorem Factorization into primes is unique up to order and associates, i.e., if $d = b_1b_2 \cdots b_n = c_1c_2 \cdots c_m$ with each b_i and each c_i prime, then $n = m$ and for some permutation σ of the indices, b_i and $c_{\sigma(i)}$ are associates for every i. Note also $\exists$ a unit u and primes $p_1, p_2, \ldots, p_t$ where no two are associates and $du = p_1^{s_1} p_2^{s_2} \cdots p_t^{s_t}$.

Proof This follows from the notes above.

Definition R is a *factorization domain* (FD) means that R is a domain and if a is a non-zero non-unit element of R, then a factors into a finite product of irreducibles.

Definition R is a *unique factorization domain* (UFD) means R is a FD in which factorization is unique (up to order and associates).

Theorem If R is a UFD and a is a non-zero non-unit of R, then a is irreducible $\Leftrightarrow$ a is prime. Thus in a UFD, elements factor as the product of primes.

Proof Suppose R is a UFD, a is an irreducible element of R, and $a|bc$. If either b or c is a unit or is zero, then a divides one of them, so suppose each of b and c is a non-zero non-unit element of R. There exists an element d with $ad = bc$. Each of b and c factors as the product of irreducibles and the product of these products is the factorization of bc. It follows from the uniqueness of the factorization of $ad = bc$, that one of these irreducibles is an associate of a, and thus $a|b$ or $a|c$. Therefore the element a is a prime.

Theorem Suppose R is a FD. Then the following are equivalent.

1) R is a UFD.
2) Every irreducible element of R is prime, i.e., a irreducible $\Leftrightarrow$ a is prime.

Proof We already know 1) $\Rightarrow$ 2). Part 2) $\Rightarrow$ 1) because factorization into primes is always unique.

This is a revealing and useful theorem. If R is a FD, then R is a UFD iff each irreducible element generates a prime ideal. Fortunately, principal ideal domains have this property, as seen in the next theorem.

Theorem Suppose R is a PID and $a \in R$ is non-zero non-unit. Then the following are equivalent.

1) aR is a maximal ideal.
2) aR is a prime ideal, i.e., a is a prime element.
3) a is irreducible.

Proof Every maximal ideal is a prime ideal, so 1) $\Rightarrow$ 2). Every prime element is an irreducible element, so 2) $\Rightarrow$ 3). Now suppose a is irreducible and show aR is a maximal ideal. If I is an ideal containing aR, $\exists\ b \in R$ with $I = bR$. Since b divides a, the element b is a unit or an associate of a. This means $I = R$ or $I = aR$.

Our goal is to prove that a PID is a UFD. Using the two theorems above, it only remains to show that a PID is a FD. The proof will not require that ideals be principally generated, but only that they be finitely generated. This turns out to be equivalent to the property that any collection of ideals has a "maximal" element. We shall see below that this is a useful concept which fits naturally into the study of unique factorization domains.

Theorem Suppose R is a commutative ring. Then the following are equivalent.

1) If $I \subset R$ is an ideal, $\exists$ a finite set $\{a_1, a_2, ..., a_n\} \subset R$ such that $I = a_1R + a_2R + \cdots + a_nR$, i.e., each ideal of R is finitely generated.
2) Any non-void collection of ideals of R contains an ideal I which is maximal in the collection. This means if J is an ideal in the collection with $J \supset I$, then $J = I$. (The ideal I is maximal only in the sense described. It need not contain all the ideals of the collection, nor need it be a maximal ideal of the ring R.)
3) If $I_1 \subset I_2 \subset I_3 \subset ...$ is a monotonic sequence of ideals, $\exists\ t_0 \geq 1$ such that $I_t = I_{t_0}$ for all $t \geq t_0$.

Proof Suppose 1) is true and show 3). The ideal $I = I_1 \cup I_2 \cup \ldots$ is finitely generated and $\exists\ t_0 \geq 1$ such that I_{t_0} contains those generators. Thus 3) is true. Now suppose 2) is true and show 1). Let I be an ideal of R, and consider the collection of all finitely generated ideals contained in I. By 2) there is a maximal one, and it must be I itself, and thus 1) is true. We now have 2)⇒1)⇒3), so suppose 2) is false and show 3) is false. So there is a collection of ideals of R such that any ideal in the collection is properly contained in another ideal of the collection. Thus it is possible to construct a sequence of ideals $I_1 \subset I_2 \subset I_3 \ldots$ with each properly contained in the next, and therefore 3) is false. (Actually this construction requires the Hausdorff Maximality Principle or some form of the Axiom of Choice, but we slide over that.)

Definition If R satisfies these properties, R is said to be *Noetherian*, or it is said to satisfy the *ascending chain condition*. This property is satisfied by many of the classical rings in mathematics. Having three definitions makes this property useful and easy to use. For example, see the next theorem.

Theorem A Noetherian domain is a FD. In particular, a PID is a FD.

Proof Suppose there is a non-zero non-unit element that does not factor as the finite product of irreducibles. Consider all ideals dR where d does not factor. Since R is Noetherian, $\exists$ a maximal one cR. The element c must be reducible, i.e., $c = ab$ where neither a nor b is a unit. Each of aR and bR properly contains cR, and so each

of a and b factors as a finite product of irreducibles. This gives a finite factorization of c into irreducibles, which is a contradiction.

Corollary A PID is a UFD. So $\mathbf{Z}$ is a UFD and if F is a field, $F[x]$ is a UFD.

You see the basic structure of UFD^s is quite easy. It takes more work to prove the following theorems, which are stated here only for reference.

Theorem If R is a UFD then $R[x_1, ..., x_n]$ is a UFD. Thus if F is a field, $F[x_1, ..., x_n]$ is a UFD. (This theorem goes all the way back to Gauss.)

If R is a PID, then the formal power series $R[[x_1, ..., x_n]]$ is a UFD. Thus if F is a field, $F[[x_1, ..., x_n]]$ is a UFD. (There is a UFD R where $R[[x]]$ is not a UFD. See page 566 of *Commutative Algebra* by N. Bourbaki.)

Theorem Germs of analytic functions on $\mathbf{C}^n$ form a UFD.

Proof See Theorem 6.6.2 of *An Introduction to Complex Analysis in Several Variables* by L. Hörmander.

Theorem Suppose R is a commutative ring. Then R is Noetherian $\Rightarrow R[x_1, ..., x_n]$ and $R[[x_1, ..., x_n]]$ are Noetherian. (This is the famous *Hilbert Basis Theorem.*)

Theorem If R is Noetherian and $I \subset R$ is a proper ideal, then R/I is Noetherian. (This follows immediately from the definition. This and the previous theorem show that Noetherian is a ubiquitous property in ring theory.)

Domains With Non-unique Factorizations Next are presented two of the standard examples of Noetherian domains that are not unique factorization domains.

Exercise Let $R = \mathbf{Z}(\sqrt{5}) = \{n + m\sqrt{5} : n, m \in \mathbf{Z}\}$. Show that R is a subring of $\mathbf{R}$ which is not a UFD. In particular $2 \cdot 2 = (1 - \sqrt{5}) \cdot (-1 - \sqrt{5})$ are two distinct irreducible factorizations of 4. Show R is isomorphic to $\mathbf{Z}[x]/(x^2 - 5)$, where $(x^2 - 5)$ represents the ideal $(x^2 - 5)\mathbf{Z}[x]$, and $R/(2)$ is isomorphic to $\mathbf{Z}_2[x]/(x^2 - [5]) = \mathbf{Z}_2[x]/(x^2 + [1])$, which is not a domain.

Exercise Let $R = \mathbf{R}[x,y,z]/(x^2 - yz)$. Show $x^2 - yz$ is irreducible and thus prime in $\mathbf{R}[x,y,z]$. If $u \in \mathbf{R}[x,y,z]$, let $\bar{u} \in R$ be the coset containing u. Show R is not a UFD. In particular $\bar{x} \cdot \bar{x} = \bar{y} \cdot \bar{z}$ are two distinct irreducible factorizations of $\bar{x}^2$. Show $R/(\bar{x})$ is isomorphic to $\mathbf{R}[y,z]/(yz)$, which is not a domain. An easier approach is to let $f : \mathbf{R}[x,y,z] \to \mathbf{R}[x,y]$ be the ring homomorphism defined by $f(x) = xy$, $f(y) = x^2$, and $f(z) = y^2$. Then $S = \mathbf{R}[xy, x^2, y^2]$ is the image of f and S is isomorphic to R. Note that xy, x^2, and y^2 are irreducible in S and $(xy)(xy) = (x^2)(y^2)$ are two distinct irreducible factorizations of $(xy)^2$ in S.

Exercise In Group Theory If G is an additive abelian group, a subgroup H of G is said to be maximal if $H \neq G$ and there are no subgroups properly between H and G. Show that H is maximal iff $G/H \approx \mathbf{Z}_p$ for some prime p. For simplicity, consider the case $G = \mathbf{Q}$. Which one of the following is true?

1) If $a \in \mathbf{Q}$, then there is a maximal subgroup H of $\mathbf{Q}$ which contains a.
2) $\mathbf{Q}$ contains no maximal subgroups.

Splitting Short Exact Sequences

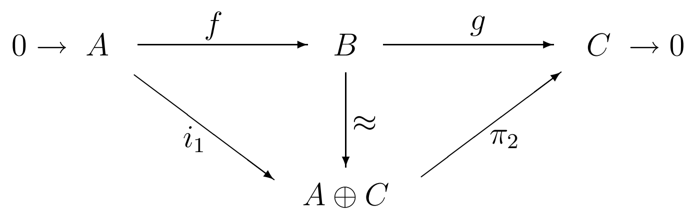
Suppose B is an R-module and K is a submodule of B. As defined in the chapter on linear algebra, K is a summand of B provided $\exists$ a submodule L of B with $K + L = B$ and $K \cap L = \underline{0}$. In this case we write $K \oplus L = B$. When is K a summand of B? It turns out that K is a summand of B iff there is a splitting map from B/K to B. In particular, if B/K is free, K must be a summand of B. This is used below to show that if R is a PID, then every submodule of R^n is free.

Theorem 1 Suppose R is a ring, B and C are R-modules, and $g : B \to C$ is a surjective homomorphism with kernel K. Then the following are equivalent.

1) K is a summand of B.
2) g has a right inverse, i.e., $\exists$ a homomorphism $h : C \to B$ with $g \circ h = I : C \to C$. (h is called a *splitting map*.)

Proof Suppose 1) is true, i.e., suppose $\exists$ a submodule L of B with $K \oplus L = B$. Then $(g|L) : L \to C$ is an isomorphism. If $i : L \to B$ is inclusion, then h defined by $h = i \circ (g|L)^{-1}$ is a right inverse of g. Now suppose 2) is true and $h : C \to B$ is a right inverse of g. Then h is injective, $K + h(C) = B$ and $K \cap h(C) = \underline{0}$. Thus $K \oplus h(C) = B$.

Definition Suppose $f : A \to B$ and $g : B \to C$ are R-module homomorphisms. The statement that $0 \to A \xrightarrow{f} B \xrightarrow{g} C \to 0$ is a *short exact sequence* (s.e.s) means f is injective, g is surjective and $f(A) = \ker(g)$. The canonical split s.e.s. is $A \to A \oplus C \to C$ where $f = i_1$ and $g = \pi_2$. A short exact sequence is said to split if $\exists$ an isomorphism $B \xrightarrow{\approx} A \oplus C$ such that the following diagram commutes.



We now restate the previous theorem in this terminology.

Theorem 1.1 A short exact sequence $0 \to A \to B \to C \to 0$ splits iff $f(A)$ is a summand of B, iff $B \to C$ has a splitting map. If C is a free R-module, there is a splitting map and thus the sequence splits.

Proof We know from the previous theorem $f(A)$ is a summand of B iff $B \to C$ has a splitting map. Showing these properties are equivalent to the splitting of the sequence is a good exercise in the art of diagram chasing. Now suppose C has a free basis $T \subset C$, and $g : B \to C$ is surjective. There exists a function $h : T \to B$ such that $g \circ h(c) = c$ for each $c \in T$. The function h extends to a homomorphism from C to B which is a right inverse of g.

Theorem 2 If R is a domain, then the following are equivalent.

1) R is a PID.
2) Every submodule of R_R is a free R-module of dimension ≤ 1.

This theorem restates the ring property of PID as a module property. Although this theorem is transparent, 1)$\Rightarrow$2) is a precursor to the following classical result.

Theorem 3 If R is a PID and $A \subset R^n$ is a submodule, then A is a free R-module of dimension $\leq n$. Thus subgroups of $\mathbf{Z}^n$ are free $\mathbf{Z}$-modules of dimension $\leq n$.

Proof From the previous theorem we know this is true for $n = 1$. Suppose $n > 1$ and the theorem is true for submodules of R^{n-1}. Suppose $A \subset R^n$ is a submodule.

Consider the following short exact sequences, where $f : R^{n-1} \to R^{n-1} \oplus R$ is inclusion and $g = \pi : R^{n-1} \oplus R \to R$ is the projection.

$$0 \longrightarrow R^{n-1} \xrightarrow{f} R^{n-1} \oplus R \xrightarrow{\pi} R \longrightarrow 0$$

$$0 \longrightarrow A \cap R^{n-1} \longrightarrow A \longrightarrow \pi(A) \longrightarrow 0$$

By induction, $A \cap R^{n-1}$ is free of dimension $\leq n-1$. If $\pi(A) = \underline{0}$, then $A \subset R^{n-1}$. If $\pi(A) \neq \underline{0}$, it is free of dimension 1 and thus the sequence splits by Theorem 1.1. In either case, A is a free submodule of dimension $\leq n$.

Exercise Let $A \subset \mathbf{Z}^2$ be the subgroup generated by $\{(6, 24), (16, 64)\}$. Show A is a free $\mathbf{Z}$-module of dimension 1. Also show the s.e.s. $\mathbf{Z}_4 \xrightarrow{\times 3} \mathbf{Z}_{12} \longrightarrow \mathbf{Z}_3$ splits but $\mathbf{Z} \xrightarrow{\times 2} \mathbf{Z} \longrightarrow \mathbf{Z}_2$ and $\mathbf{Z}_2 \xrightarrow{\times 2} \mathbf{Z}_4 \longrightarrow \mathbf{Z}_2$ do not (see top of page 78).

Euclidean Domains

The ring $\mathbf{Z}$ possesses the Euclidean algorithm and the polynomial ring $F[x]$ has the division algorithm (pages 14 and 45). The concept of *Euclidean domain* is an abstraction of these properties, and the efficiency of this abstraction is displayed in this section. Furthermore the first axiom, $\phi(a) \leq \phi(ab)$, is used only in Theorem 2, and is sometimes omitted from the definition. Anyway it is possible to just play around with matrices and get some deep results. If R is a Euclidean domain and M is a finitely generated R-module, then M is the sum of cyclic modules. This is one of the great classical theorems of abstract algebra, and you don't have to worry about it becoming obsolete. Here $\mathbf{N}$ will denote the set of all non-negative integers, not just the set of positive integers.

Definition A domain R is a *Euclidean domain* provided $\exists\ \phi : (R - \underline{0}) \longrightarrow \mathbf{N}$ such that if $a, b \in (R - \underline{0})$, then

1) $\phi(a) \leq \phi(ab)$.
2) $\exists\ q, r \in R$ such that $a = bq + r$ with $r = \underline{0}$ or $\phi(r) < \phi(b)$.

Examples of Euclidean Domains

$\mathbf{Z}$ with $\phi(n) = |n|$.
A field F with $\phi(a) = 1\ \forall\ a \neq \underline{0}$ or with $\phi(a) = 0\ \forall\ a \neq \underline{0}$.
$F[x]$ where F is a field with $\phi(f = a_0 + a_1 x + \cdots + a_n x^n) =$ deg(f).
$\mathbf{Z}[i] = \{a + bi : a, b \in \mathbf{Z}\}$ = Gaussian integers with $\phi(a + bi) = a^2 + b^2$.

Theorem 1 If R is a Euclidean domain, then R is a PID and thus a UFD.

Proof If I is a non-zero ideal, then $\exists\ b \in I - \underline{0}$ satisfying $\phi(b) \leq \phi(a)\ \forall\ a \in I - \underline{0}$. Then b generates I because if $a \in I - \underline{0}$, $\exists\ q, r$ with $a = bq + r$. Now $r \in I$ and $r \neq \underline{0} \Rightarrow \phi(r) < \phi(b)$ which is impossible. Thus $r = \underline{0}$ and $a \in bR$ so $I = bR$.

Theorem 2 If R is a Euclidean domain and $a, b \in R - \underline{0}$, then

$\phi(\underline{1})$ is the smallest integer in the image of ϕ.
a is a unit in R iff $\phi(a) = \phi(\underline{1})$.
a and b are associates $\Rightarrow$ $\phi(a) = \phi(b)$.

Proof This is a good exercise. However it is unnecessary for Theorem 3 below.

The following remarkable theorem is the foundation for the results of this section.

Theorem 3 If R is a Euclidean domain and $(a_{i,j}) \in R_{n,t}$ is a non-zero matrix, then by elementary row and column operations $(a_{i,j})$ can be transformed to

$$\begin{pmatrix} d_1 & 0 & \cdots & & & 0 \\ 0 & d_2 & & & & \\ \vdots & & \ddots & & & \\ & & & d_m & & \\ & & & & 0 & \\ & & & & & \\ 0 & & & & & 0 \end{pmatrix}$$

where each $d_i \neq \underline{0}$, and $d_i | d_{i+1}$ for $1 \leq i < m$. Also d_1 generates the ideal of R generated by the entries of $(a_{i,j})$.

Proof Let $I \subset R$ be the ideal generated by the elements of the matrix $A = (a_{i,j})$. If $E \in R_n$, then the ideal J generated by the elements of EA has $J \subset I$. If E is invertible, then $J = I$. In the same manner, if $E \in R_t$ is invertible and J is the ideal generated by the elements of AE, then $J = I$. This means that row and column operations on A do not change the ideal I. Since R is a PID, there is an element d_1 with $I = d_1 R$, and this will turn out to be the d_1 displayed in the theorem.

The matrix $(a_{i,j})$ has at least one non-zero element d with $\phi(d)$ a miminum. However, row and column operations on $(a_{i,j})$ may produce elements with smaller

ϕ values. To consolidate this approach, consider matrices obtained from $(a_{i,j})$ by a finite number of row and column operations. Among these, let $(b_{i,j})$ be one which has an entry $d_1 \neq 0$ with $\phi(d_1)$ a minimum. By elementary operations of type 2, the entry d_1 may be moved to the $(1,1)$ place in the matrix. Then d_1 will divide the other entries in the first row, else we could obtain an entry with a smaller ϕ value. Thus by column operations of type 3, the other entries of the first row may be made zero. In a similar manner, by row operations of type 3, the matrix may be changed to the following form.

$$\begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & c_{ij} & \\ 0 & & & \end{pmatrix}$$

Note that d_1 divides each $c_{i,j}$, and thus $I = d_1 R$. The proof now follows by induction on the size of the matrix.

This is an example of a theorem that is easy to prove playing around at the blackboard. Yet it must be a deep theorem because the next two theorems are easy consequences.

Theorem 4 Suppose R is a Euclidean domain, B is a finitely generated free R-module and $A \subset B$ is a non-zero submodule. Then $\exists$ free bases $\{a_1, a_2, ..., a_t\}$ for A and $\{b_1, b_2, ..., b_n\}$ for B, with $t \leq n$, and such that each $a_i = d_i b_i$, where each $d_i \neq \underline{0}$, and $d_i | d_{i+1}$ for $1 \leq i < t$. Thus $B/A \approx R/d_1 \oplus R/d_2 \oplus \cdots \oplus R/d_t \oplus R^{n-t}$.

Proof By Theorem 3 in the section Splitting Short Exact Sequences, A has a free basis $\{v_1, v_2, ..., v_t\}$. Let $\{w_1, w_2, ..., w_n\}$ be a free basis for B, where $n \geq t$. The composition

$$R^t \xrightarrow{\approx} A \xrightarrow{\subset} B \xrightarrow{\approx} R^n$$

$$e_i \longrightarrow v_i \qquad w_i \longrightarrow e_i$$

is represented by a matrix $(a_{i,j}) \in R_{n,t}$ where $v_i = a_{1,i}w_1 + a_{2,i}w_2 + \cdots + a_{n,i}w_n$. By the previous theorem, $\exists$ invertible matrixes $U \in R_n$ and $V \in R_t$ such that

$$U(a_{i,j})V = \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & 0 & \\ \vdots & 0 & \ddots & \\ & & & d_t \\ & & & \\ 0 & \cdots & & 0 \end{pmatrix}$$

with $d_i | d_{i+1}$. Since changing the isomorphisms $R^t \xrightarrow{\approx} A$ and $B \xrightarrow{\approx} R^n$ corresponds to changing the bases $\{v_1, v_2, ..., v_t\}$ and $\{w_1, w_2, ..., w_n\}$, the theorem follows.

Theorem 5 If R is a Euclidean domain and M is a finitely generated R-module, then $M \approx R/d_1 \oplus R/d_2 \oplus \cdots \oplus R/d_t \oplus R^m$ where each $d_i \neq \underline{0}$, and $d_i | d_{i+1}$ for $1 \leq i < t$.

Proof By hypothesis $\exists$ a finitely generated free module B and a surjective homomorphism $B \longrightarrow M \longrightarrow 0$. Let A be the kernel, so $0 \longrightarrow A \xrightarrow{\subset} B \longrightarrow M \longrightarrow 0$ is a s.e.s. and $B/A \approx M$. The result now follows from the previous theorem.

The way Theorem 5 is stated, some or all of the elements d_i may be units, and for such d_i, $R/d_i = \underline{0}$. If we assume that no d_i is a unit, then the elements $d_1, d_2, ..., d_t$ are called *invariant factors*. They are unique up to associates, but we do not bother with that here. If $R = \mathbf{Z}$ and we select the d_i to be positive, they are unique. If $R = F[x]$ and we select the d_i to be monic, then they are unique. The splitting in Theorem 5 is not the ultimate because the modules R/d_i may split into the sum of other cyclic modules. To prove this we need the following Lemma.

Lemma Suppose R is a PID and b and c are non-zero non-unit elements of R. Suppose b and c are relatively prime, i.e., there is no prime common to their prime factorizations. Then bR and cR are comaximal ideals. (See p 108 for comaximal.)

Proof There exists an $a \in R$ with $aR = bR + cR$. Since $a|b$ and $a|c$, a is a unit, so $R = bR + cR$.

Theorem 6 Suppose R is a PID and d is a non-zero non-unit element of R. Assume $d = p_1^{s_1} p_2^{s_2} \cdots p_t^{s_t}$ is the prime factorization of d (see bottom of p 110). Then the natural map $R/d \xrightarrow{\approx} R/p_1^{s_1} \oplus \cdots \oplus R/p_t^{s_t}$ is an isomorphism of R-modules. (The elements $p_i^{s_i}$ are called *elementary divisors* of R/d.)

Proof If $i \neq j$, $p_i^{s_i}$ and $p_j^{s_j}$ are relatively prime. By the Lemma above, they are

comaximal and thus by the Chinese Remainder Theorem, the natural map is a ring isomorphism (page 108). Since the natural map is also an R-module homomorphism, it is an R-module isomorphism.

This theorem carries the splitting as far as it can go, as seen by the next exercise.

Exercise Suppose R is a PID, $p \in R$ is a prime element, and $s \geq 1$. Then the R-module R/p^s has no proper submodule which is a summand.

Torsion Submodules This will give a little more perspective to this section.

Definition Suppose M is a module over a domain R. An element $m \in M$ is said to be a *torsion element* if $\exists\ r \in R$ with $r \neq \underline{0}$ and $mr = \underline{0}$. This is the same as saying m is dependent. If $R = \mathbf{Z}$, it is the same as saying m has finite order. Denote by $T(M)$ the set of all torsion elements of M. If $T(M) = \underline{0}$, we say that M is torsion free.

Theorem 7 Suppose M is a module over a domain R. Then $T(M)$ is a submodule of M and $M/T(M)$ is torsion free.

Proof This is a simple exercise.

Theorem 8 Suppose R is a Euclidean domain and M is a finitely generated R-module which is torsion free. Then M is a free R-module, i.e., $M \approx R^m$.

Proof This follows immediately from Theorem 5.

Theorem 9 Suppose R is a Euclidean domain and M is a finitely generated R-module. Then the following s.e.s. splits.

$$0 \longrightarrow T(M) \longrightarrow M \longrightarrow M/T(M) \longrightarrow 0$$

Proof By Theorem 7, $M/T(M)$ is torsion free. By Theorem 8, $M/T(M)$ is a free R-module, and thus there is a splitting map. Of course this theorem is transparent anyway, because Theorem 5 gives a splitting of M into a torsion part and a free part.

Note It follows from Theorem 9 that $\exists$ a free submodule V of M such that $T(M) \oplus V = M$. The first summand $T(M)$ is unique, but the complementary summand V is not unique. V depends upon the splitting map and is unique only up to isomorphism.

To complete this section, here are two more theorems that follow from the work we have done.

Theorem 10 Suppose T is a domain and T^* is the multiplicative group of units of T. If G is a finite subgroup of T^*, then G is a cyclic group. Thus if F is a finite field, the multiplicative group F^* is cyclic. Thus if p is a prime, $(\mathbf{Z}_p)^*$ is cyclic.

Proof This is a corollary to Theorem 5 with $R = \mathbf{Z}$. The multiplicative group G is isomorphic to an additive group $\mathbf{Z}/d_1 \oplus \mathbf{Z}/d_2 \oplus \cdots \oplus \mathbf{Z}/d_t$ where each $d_i > 1$ and $d_i | d_{i+1}$ for $1 \leq i < t$. Every u in the additive group has the property that $ud_t = \underline{0}$. So every $g \in G$ is a solution to $x^{d_t} - \underline{1} = \underline{0}$. If $t > 1$, the equation will have degree less than the number of roots, which is impossible. Thus $t = 1$ and so G is cyclic.

Exercise For which primes p and q is the group of units $(\mathbf{Z}_p \times \mathbf{Z}_q)^*$ a cyclic group?

We know from Exercise 2) on page 59 that an invertible matrix over a field is the product of elementary matrices. This result also holds for any invertible matrix over a Euclidean domain.

Theorem 11 Suppose R is a Euclidean domain and $A \in R_n$ is a matrix with non-zero determinant. Then by elementary row and column operations, A may be transformed to a diagonal matrix

$$\begin{pmatrix} d_1 & & & 0 \\ & d_2 & & \\ & & \ddots & \\ 0 & & & d_n \end{pmatrix}$$

where each $d_i \neq \underline{0}$ and $d_i | d_{i+1}$ for $1 \leq i < n$. Also d_1 generates the ideal generated by the entries of A. Furthermore A is invertible iff each d_i is a unit. Thus if A is invertible, A is the product of elementary matrices.

Proof It follows from Theorem 3 that A may be transformed to a diagonal matrix with $d_i|d_{i+1}$. Since the determinant of A is not zero, it follows that each $d_i \neq \underline{0}$. Furthermore, the matrix A is invertible iff the diagonal matrix is invertible, which is true iff each d_i is a unit. If each d_i is a unit, then the diagonal matrix is the product of elementary matrices of type 1. Therefore if A is invertible, it is the product of elementary matrices.

Exercise Let $R = \mathbf{Z}$, $A = \begin{pmatrix} 3 & 11 \\ 0 & 4 \end{pmatrix}$ and $D = \begin{pmatrix} 3 & 11 \\ 1 & 4 \end{pmatrix}$. Perform elementary operations on A and D to obtain diagonal matrices where the first diagonal element divides the second diagonal element. Write D as the product of elementary matrices. Find the characteristic polynomials of A and D. Find an elementary matrix B over $\mathbf{Z}$ such that $B^{-1}AB$ is diagonal. Find an invertible matrix C in $\mathbf{R}_2$ such that $C^{-1}DC$ is diagonal. Show C cannot be selected in $\mathbf{Q}_2$.

Jordan Blocks

In this section, we define the two special types of square matrices used in the Rational and Jordan canonical forms. Note that the Jordan block $B(q)$ is the sum of a scalar matrix and a nilpotent matrix. A Jordan block displays its eigenvalue on the diagonal, and is more interesting than the companion matrix $C(q)$. But as we shall see later, the Rational canonical form will always exist, while the Jordan canonical form will exist iff the characteristic polynomial factors as the product of linear polynomials.

Suppose R is a commutative ring, $q = a_0 + a_1x + \cdots + a_{n-1}x^{n-1} + x^n \in R[x]$ is a monic polynomial of degree $n \geq 1$, and V is the $R[x]$-module $V = R[x]/q$. V is a torsion module over the ring $R[x]$, but as an R-module, V has a free basis $\{1, x, x^2, \ldots, x^{n-1}\}$. (See the last part of the last theorem on page 46.) Multiplication by x defines an R-module endomorphism on V, and $C(q)$ will be the matrix of this endomorphism with respect to this basis. Let $T : V \to V$ be defined by $T(v) = vx$. If $h(x) \in R[x]$, $h(T)$ is the R-module homomorphism given by multiplication by $h(x)$. The homomorphism from $R[x]/q$ to $R[x]/q$ given by multiplication by $h(x)$, is zero iff $h(x) \in qR[x]$. That is to say $q(T) = a_0I + a_1T + \cdots + T^n$ is the zero homomorphism, and $h(T)$ is the zero homomorphism iff $h(x) \in qR[x]$. All of this is supposed to make the next theorem transparent.

Theorem Let V have the free basis $\{1, x, x^2, ..., x^{n-1}\}$. The companion matrix

representing T is

$$C(q) = \begin{pmatrix} 0 & \ldots & \ldots & 0 & -a_0 \\ 1 & 0 & \ldots & 0 & -a_1 \\ 0 & 1 & 0 & & -a_2 \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & \ldots & \ldots & 1 & -a_{n-1} \end{pmatrix}$$

The characteristic polynomial of $C(q)$ is q, and $|C(q)| = (-1)^n a_0$. Finally, if $h(x) \in R[x]$, $h(C(q))$ is zero iff $h(x) \in qR[x]$.

Theorem Suppose $\lambda \in R$ and $q(x) = (x - \lambda)^n$. Let V have the free basis $\{1, (x-\lambda), (x-\lambda)^2, \ldots, (x-\lambda)^{n-1}\}$. Then the matrix representing T is

$$B(q) = \begin{pmatrix} \lambda & 0 & \ldots & \ldots & 0 \\ 1 & \lambda & 0 & \ldots & 0 \\ 0 & 1 & \lambda & & \vdots \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & \ldots & \ldots & 1 & \lambda \end{pmatrix}$$

The characteristic polynomial of $B(q)$ is q, and $|B(q)| = \lambda^n = (-1)^n a_0$. Finally, if $h(x) \in R[x]$, $h(B(q))$ is zero iff $h(x) \in qR[x]$.

Note For $n = 1$, $C(a_0 + x) = B(a_0 + x) = (-a_0)$. This is the only case where a block matrix may be the zero matrix.

Note In $B(q)$, if you wish to have the 1^s above the diagonal, reverse the order of the basis for V.

Jordan Canonical Form

We are finally ready to prove the Rational and Jordan forms. Using the previous sections, all that's left to do is to put the pieces together. (For an overview of Jordan form, read first the section in Chapter 5, page 96.)

Suppose R is a commutative ring, V is an R-module, and $T : V \to V$ is an R-module homomorphism. Define a scalar multiplication $V \times R[x] \to V$ by $v(a_0 + a_1x + \cdots + a_rx^r) = va_0 + T(v)a_1 + \cdots + T^r(v)a_r$.

Theorem 1 Under this scalar multiplication, V is an $R[x]$-module.

This is just an observation, but it is one of the great tricks in mathematics. Questions about the transformation T are transferred to questions about the module V over the ring $R[x]$. And in the case R is a field, $R[x]$ is a Euclidean domain and so we know almost everything about V as an $R[x]$-module.

Now in this section, we suppose R is a field F, V is a finitely generated F-module, $T : V \to V$ is a linear transformation and V is an $F[x]$-module with $vx = T(v)$. Our goal is to select a basis for V such that the matrix representing T is in some simple form. A submodule of $V_{F[x]}$ is a submodule of V_F which is invariant under T. We know $V_{F[x]}$ is the sum of cyclic modules from Theorems 5 and 6 in the section on Euclidean Domains. Since V is finitely generated as an F-module, the free part of this decomposition will be zero. In the section on Jordan Blocks, a basis is selected for these cyclic modules and the matrix representing T is described. This gives the Rational Canonical Form and that is all there is to it. If all the eigenvalues for T are in F, we pick another basis for each of the cyclic modules (see the second theorem in the section on Jordan Blocks). Then the matrix representing T is called the Jordan Canonical Form. Now we say all this again with a little more detail.

From Theorem 5 in the section on Euclidean Domains, it follows that

$$V_{F[x]} \approx F[x]/d_1 \oplus F[x]/d_2 \oplus \cdots \oplus F[x]/d_t$$

where each d_i is a monic polynomial of degree ≥ 1, and $d_i | d_{i+1}$. Pick $\{1, x, x^2, \ldots, x^{m-1}\}$ as the F-basis for $F[x]/d_i$ where m is the degree of the polynomial d_i.

Theorem 2 With respect to this basis, the matrix representing T is

$$\begin{pmatrix} C(d_1) & & & \\ & C(d_2) & & \\ & & \ddots & \\ & & & C(d_t) \end{pmatrix}$$

The characteristic polynomial of T is $p = d_1 d_2 \cdots d_t$ and $p(T) = \underline{0}$. This is a type of canonical form but it does not seem to have a name.

Now we apply Theorem 6 to each $F[x]/d_i$. This gives $V_{F[x]} \approx F[x]/p_1^{s_1} \oplus \cdots \oplus F[x]/p_r^{s_r}$ where the p_i are irreducible monic polynomials of degree at least 1. The p_i need not be distinct. Pick an F-basis for each $F[x]/p_i^{s_i}$ as before.

Theorem 3 With respect to this basis, the matrix representing T is

$$\begin{pmatrix} C(p_1^{s_1}) & & & 0 \\ & C(p_2^{s_2}) & & \\ & & \ddots & \\ 0 & & & C(p_r^{s_r}) \end{pmatrix}$$

The characteristic polynomial of T is $p = p_1^{s_1} \cdots p_r^{s_r}$ and $p(T) = \underline{0}$. This is called the *Rational canonical form* for T.

Now suppose the characteristic polynomial of T factors in $F[x]$ as the product of linear polynomials. Thus in the Theorem above, $p_i = x - \lambda_i$ and

$$V_{F[x]} \approx F[x]/(x-\lambda_1)^{s_1} \oplus \cdots \oplus F[x]/(x-\lambda_r)^{s_r}$$

is an isomorphism of $F[x]$-modules. Pick $\{1, (x-\lambda_i), (x-\lambda_i)^2, \ldots, (x-\lambda_i)^{m-1}\}$ as the F-basis for $F[x]/(x-\lambda_i)^{s_i}$ where m is s_i.

Theorem 4 With respect to this basis, the matrix representing T is

$$\begin{pmatrix} B((x-\lambda_1)^{s_1}) & & & 0 \\ & B((x-\lambda_2)^{s_2}) & & \\ 0 & & \ddots & \\ & & & B((x-\lambda_r)^{s_r}) \end{pmatrix}$$

The characteristic polynomial of T is $p = (x-\lambda_1)^{s_1}\cdots(x-\lambda_r)^{s_r}$ and $p(T) = \underline{0}$. This is called the *Jordan canonical form* for T. Note that the λ_i need not be distinct.

Note A diagonal matrix is in Rational canonical form and in Jordan canonical form. This is the case where each block is one by one. Of course a diagonal matrix is about as canonical as you can get. Note also that if a matrix is in Jordan form, its trace is the sum of the eigenvalues and its determinant is the product of the eigenvalues. Finally, this section is loosely written, so it is important to use the transpose principle to write three other versions of the last two theorems.

Exercise Suppose F is a field of characteristic 0 and $T \in F_n$ has trace$(T^i) = \underline{0}$ for $0 < i \leq n$. Show T is nilpotent. Let $p \in F[x]$ be the characteristic polynomial of T. The polynomial p may not factor into linears in $F[x]$, and thus T may have no conjugate in F_n which is in Jordan form. However this exercise can still be worked using Jordan form. This is based on the fact that there exists a field $\bar{F}$ containing F as a subfield, such that p factors into linears in $\bar{F}[x]$. This fact is not proved in this book, but it is assumed for this exercise. So $\exists$ an invertible matrix $U \in \bar{F}_n$ so that $U^{-1}TU$ is in Jordan form, and of course, T is nilpotent iff $U^{-1}TU$ is nilpotent. The point is that it sufficies to consider the case where T is in Jordan form, and to show the diagonal elements are all zero.

So suppose T is in Jordan form and trace $(T^i) = \underline{0}$ for $1 \leq i \leq n$. Thus trace $(p(T)) = a_0 n$ where a_0 is the constant term of $p(x)$. We know $p(T) = \underline{0}$ and thus trace $(p(T)) = \underline{0}$, and thus $a_0 n = \underline{0}$. Since the field has characteristic 0, $a_0 = \underline{0}$ and so $\underline{0}$ is an eigenvalue of T. This means that one block of T is a strictly lower triangular matrix. Removing this block leaves a smaller matrix which still satisfies the hypothesis, and the result follows by induction on the size of T. This exercise illustrates the power and facility of Jordan form. It also has a cute corollary.

Corollary Suppose F is a field of characteristic 0, $n \geq 1$, and $(\lambda_1, \lambda_2, .., \lambda_n) \in F^n$ satisfies $\lambda_1^i + \lambda_2^i + \cdots + \lambda_n^i = \underline{0}$ for each $1 \leq i \leq n$. Then $\lambda_i = \underline{0}$ for $1 \leq i \leq n$.

Minimal polynomials To conclude this section here are a few comments on the minimal polynomial of a linear transformation. This part should be studied only if you need it. Suppose V is an n-dimensional vector space over a field F and $T : V \to V$ is a linear transformation. As before we make V a module over $F[x]$ with $T(v) = vx$.

Definition $Ann(V_{F[x]})$ is the set of all $h \in F[x]$ which annihilate V, i.e., which satisfy $Vh = \underline{0}$. This is a non-zero ideal of $F[x]$ and is thus generated by a unique monic polynomial $u(x) \in F(x)$, $Ann(V_{F[x]}) = uF[x]$. The polynomial u is called the *minimal polynomial* of T. Note that $u(T) = \underline{0}$ and if $h(x) \in F[x]$, $h(T) = \underline{0}$ iff h is a multiple of u in $F[x]$. If $p(x) \in F[x]$ is the characteristic polynomial of T, $p(T) = \underline{0}$ and thus p is a multiple of u.

Now we state this again in terms of matrices. Suppose $A \in F_n$ is a matrix representing T. Then $u(A) = \underline{0}$ and if $h(x) \in F[x]$, $h(A) = \underline{0}$ iff h is a multiple of u in $F[x]$. If $p(x) \in F[x]$ is the characteristic polynomial of A, then $p(A) = \underline{0}$ and thus p is a multiple of u. The polynomial u is also called the minimal polynomial of A. Note that these properties hold for any matrix representing T, and thus similar matrices have the same minimal polynomial. If A is given to start with, use the linear transformation $T : F^n \to F^n$ determined by A to define the polynomial u.

Now suppose $q \in F[x]$ is a monic polynomial and $C(q) \in F_n$ is the companion matrix defined in the section Jordan Blocks. Whenever $q(x) = (x - \lambda)^n$, let $B(q) \in F_n$ be the Jordan block matrix also defined in that section. Recall that q is the characteristic polynomial and the minimal polynomial of each of these matrices. This together with the rational form and the Jordan form will allow us to understand the relation of the minimal polynomial to the characteristic polynomial.

Exercise Suppose $A_i \in F_{n_i}$ has q_i as its characteristic polynomial and its minimal polynomial, and $A = \begin{pmatrix} A_1 & & & 0 \\ & A_2 & & \\ & & \ddots & \\ 0 & & & A_r \end{pmatrix}$. Find the characteristic polynomial and the minimal polynomial of A.

Exercise Suppose $A \in F_n$.

1) Suppose A is the matrix displayed in Theorem 2 above. Find the characteristic and minimal polynomials of A.

2) Suppose A is the matrix displayed in Theorem 3 above. Find the characteristic and minimal polynomials of A.

3) Suppose A is the matrix displayed in Theorem 4 above. Find the characteristic and minimal polynomials of A.

4) Suppose $\lambda \in F$. Show λ is a root of the characteristic polynomial of A iff λ is a root of the minimal polynomial of A. Show that if λ is a root, its order in the characteristic polynomial is at least as large as its order in the minimal polynomial.

5) Suppose $\bar{F}$ is a field containing F as a subfield. Show that the minimal polynomial of $A \in F_n$ is the same as the minimal polynomial of A considered as a matrix in $\bar{F}_n$. (This funny looking exercise is a little delicate.)

6) Let $F = \mathbf{R}$ and $A = \begin{pmatrix} 5 & -1 & 3 \\ 0 & 2 & 0 \\ -3 & 1 & -1 \end{pmatrix}$. Find the characteristic and minimal polynomials of A.

Determinants

In the chapter on matrices, it is stated without proof that the determinant of the product is the product of the determinants (see page 63). The purpose of this section is to give a proof of this. We suppose R is a commutative ring, C is an R-module, $n \geq 2$, and $B_1, B_2, \ldots, B_n$ is a sequence of R-modules.

Definition A map $f : B_1 \oplus B_2 \oplus \cdots \oplus B_n \to C$ is R-*multilinear* means that if $1 \leq i \leq n$, and $b_j \in B_j$ for $j \neq i$, then $f|(b_1, b_2, \ldots, B_i, \ldots, b_n)$ defines an R-linear map from B_i to C.

Theorem The set of all R-multilinear maps is an R-module.

Proof From the first exercise in Chapter 5, the set of all functions from $B_1 \oplus B_2 \oplus \cdots \oplus B_n$ to C is an R-module (see page 69). It must be seen that the R-multilinear maps form a submodule. It is easy to see that if f_1 and f_2 are R-multilinear, so is $f_1 + f_2$. Also if f is R-multilinear and $r \in R$, then (fr) is R-multilinear.

From here on, suppose $B_1 = B_2 = \cdots = B_n = B$.

Definition

1) f is *symmetric* means $f(b_1, \ldots, b_n) = f(b_{\tau(1)}, \ldots, b_{\tau(n)})$ for all permutations τ on $\{1, 2, \ldots, n\}$.
2) f is *skew-symmetric* if $f(b_1, \ldots, b_n) = \text{sign}(\tau) f(b_{\tau(1)}, \ldots, b_{\tau(n)})$ for all τ.

3) f is *alternating* if $f(b_1, \ldots, b_n) = \underline{0}$ whenever some $b_i = b_j$ for $i \neq j$.

Theorem

i) Each of these three types defines a submodule of the set of all R-multilinear maps.
ii) Alternating $\Rightarrow$ skew-symmetric.
iii) If no element of C has order 2, then alternating $\Longleftrightarrow$ skew-symmetric.

Proof Part i) is immediate. To prove ii), assume f is alternating. It sufficies to show that $f(b_1, ..., b_n) = -f(b_{\tau(1)}, ..., b_{\tau(n)})$ where τ is a transposition. For simplicity, assume $\tau = (1, 2)$. Then $\underline{0} = f(b_1 + b_2, b_1 + b_2, b_3, ..., b_n) = f(b_1, b_2, b_3, ..., b_n) + f(b_2, b_1, b_3, ..., b_n)$ and the result follows. To prove iii), suppose f is skew symmetric and no element of C has order 2, and show f is alternating. Suppose for convenience that $b_1 = b_2$ and show $f(b_1, b_1, b_3, \ldots, b_n) = \underline{0}$. If we let τ be the transposition $(1, 2)$, we get $f(b_1, b_1, b_3, \ldots, b_n) = -f(b_1, b_1, b_3, \ldots, b_n)$, and so $2f(b_1, b_1, b_3, \ldots, b_n) = \underline{0}$, and the result follows.

Now we are ready for determinant. Suppose $C = R$. In this case multilinear maps are usually called *multilinear forms*. Suppose B is R^n with the canonical basis $\{e_1, e_2, \ldots, e_n\}$. (We think of a matrix $A \in R_n$ as n column vectors, i.e., as an element of $B \oplus B \oplus \cdots \oplus B$.) First we recall the definition of determinant.

Suppose $A = (a_{i,j}) \in R_n$. Define $d : B \oplus B \oplus \cdots \oplus B \to R$ by $d(a_{1,1}e_1 + a_{2,1}e_2 + \cdots + a_{n,1}e_n, \ldots.., a_{1,n}e_1 + a_{2,n}e_2 + \cdots + a_{n,n}e_n) = \sum_{\text{all } \tau} \text{sign}(\tau)(a_{\tau(1),1}a_{\tau(2),2} \cdots a_{\tau(n),n}) = |A|$.

The next theorem follows from the section on determinants on page 61.

Theorem d is an alternating multilinear form with $d(e_1, e_2, \ldots, e_n) = \underline{1}$.

If $c \in R$, dc is an alternating multilinear form, because the set of alternating forms is an R-module. It turns out that this is all of them, as seen by the following theorem.

Theorem Suppose $f : B \oplus B \oplus \ldots \oplus B \to R$ is an alternating multilinear form. Then $f = df(e_1, e_2, \ldots, e_n)$. This means f is the multilinear form d times the scalar $f(e_1, e_2, ..., e_n)$. In other words, if $A = (a_{i,j}) \in R_n$, then $f(a_{1,1}e_1 + a_{2,1}e_2 + \cdots + a_{n,1}e_n, \ldots.., a_{1,n}e_2 + a_{2,n}e_2 + \cdots + a_{n,n}e_n) = |A|f(e_1, e_2, ..., e_n)$. Thus the set of alternating forms is a free R-module of dimension 1, and the determinant is a generator.

Proof For $n = 2$, you can simply write it out. $f(a_{1,1}e_1 + a_{2,1}e_2, a_{1,2}e_1 + a_{2,2}e_2) = a_{1,1}a_{1,2}f(e_1, e_1) + a_{1,1}a_{2,2}f(e_1, e_2) + a_{2,1}a_{1,2}f(e_2, e_1) + a_{2,1}a_{2,2}f(e_2, e_2) = (a_{1,1}a_{2,2} - a_{1,2}a_{2,1})f(e_1, e_2) = |A|f(e_1, e_2)$. For the general case, $f(a_{1,1}e_1 + a_{2,1}e_2 + \cdots + a_{n,1}e_n, ..., a_{1,n}e_1 + a_{2,n}e_2 + \cdots + a_{n,n}e_n) = \sum a_{i_1,1}a_{i_2,2} \cdots a_{i_n,n}f(e_{i_1}, e_{i_2}, ..., e_{i_n})$ where the sum is over all $1 \leq i_1 \leq n,\ 1 \leq i_2 \leq n, ..., 1 \leq i_n \leq n$. However, if any $i_s = i_t$ for $s \neq t$, that term is 0 because f is alternating. Therefore the sum is just $\sum_{\text{all } \tau} a_{\tau(1),1}a_{\tau(2),2} \cdots a_{\tau(n),n}f(e_{\tau(1)}, e_{\tau(2)}, \ldots, e_{\tau(n)}) = \sum_{\text{all } \tau} \text{sign}(\tau)a_{\tau(1),1} a_{\tau(2),2} \cdots a_{\tau(n),n}f(e_1, e_2, \ldots, e_n) = |A|f(e_1, e_2, ..., e_n)$.

This incredible classification of these alternating forms makes the proof of the following theorem easy. (See the third theorem on page 63.)

Theorem If $C, A \in R_n$, then $|CA| = |C||A|$.

Proof Suppose $C \in R_n$. Define $f : R_n \to R$ by $f(A) = |CA|$. In the notation of the previous theorem, $B = R^n$ and $R_n = R^n \oplus R^n \oplus \cdots \oplus R^n$. If $A \in R_n, A = (A_1, A_2, ..., A_n)$ where $A_i \in R^n$ is column i of A, and $f : R^n \oplus \cdots \oplus R^n \to R$ has $f(A_1, A_2, ..., A_n) = |CA|$. Use the fact that $CA = (CA_1, CA_2, ..., CA_n)$ to show that f is an alternating multilinear form. By the previous theorem, $f(A) = |A|f(e_1, e_2, ..., e_n)$. Since $f(e_1, e_2, ..., e_n) = |CI| = |C|$, it follows that $|CA| = f(A) = |A||C|$.

Dual Spaces

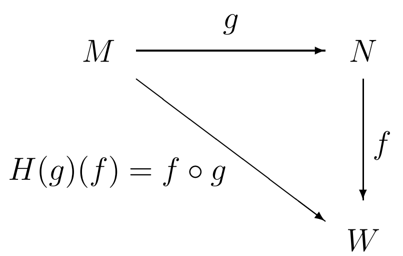
The concept of dual module is basic, not only in algebra, but also in other areas such as differential geometry and topology. If V is a finitely generated vector space over a field F, its dual V^* is defined as $V^* = \text{Hom}_F(V, F)$. V^* is isomorphic to V, but in general there is no natural isomorphism from V to V^*. However there is a natural isomorphism from V to V^{**}, and so V^* is the dual of V and V may be considered to be the dual of V^*. This remarkable fact has many expressions in mathematics. For example, a tangent plane to a differentiable manifold is a real vector space. The union of these spaces is the tangent bundle, while the union of the dual spaces is the cotangent bundle. Thus the tangent (cotangent) bundle may be considered to be the dual of the cotangent (tangent) bundle. The sections of the tangent bundle are called vector fields while the sections of the cotangent bundle are called 1-forms.

In algebraic topology, homology groups are derived from chain complexes, while cohomology groups are derived from the dual chain complexes. The sum of the cohomology groups forms a ring, while the sum of the homology groups does not.

Thus the concept of dual module has considerable power. We develop here the basic theory of dual modules.

Suppose R is a commutative ring and W is an R-module.

Definition If M is an R-module, let $H(M)$ be the R-module $H(M)=\text{Hom}_R(M, W)$. If M and N are R-modules and $g : M \to N$ is an R-module homomorphism, let $H(g) : H(N) \to H(M)$ be defined by $H(g)(f) = f \circ g$. Note that $H(g)$ is an R-module homomorphism.

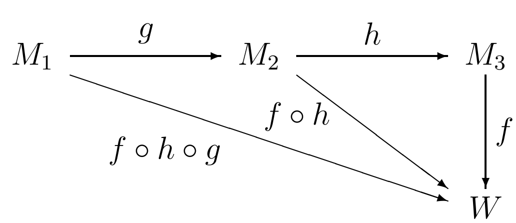


Theorem

i) If M_1 and M_2 are R-modules, $H(M_1 \oplus M_2) \approx H(M_1) \oplus H(M_2)$.

ii) If $I : M \to M$ is the identity, then $H(I) : H(M) \to H(M)$ is the identity.

iii) If $M_1 \xrightarrow{g} M_2 \xrightarrow{h} M_3$ are R-module homomorphisms, then $H(g) \circ H(h) = H(h \circ g)$. If $f : M_3 \to W$ is a homomorphism, then $(H(g) \circ H(h))(f) = H(h \circ g)(f) = f \circ h \circ g$.



Note In the language of the category theory, H is a contravariant functor from the category of R-modules to itself.

Theorem If M and N are R-modules and $g : M \to N$ is an isomorphism, then $H(g) : H(N) \to H(M)$ is an isomorphism with $H(g^{-1}) = H(g)^{-1}$.

Proof

$$I_{H(N)} = H(I_N) = H(g \circ g^{-1}) = H(g^{-1}) \circ H(g)$$

$$I_{H(M)} = H(I_M) = H(g^{-1} \circ g) = H(g) \circ H(g^{-1})$$

Theorem

i) If $g : M \to N$ is a surjective homomorphism, then $H(g) : H(N) \to H(M)$ is injective.

ii) If $g : M \to N$ is an injective homomorphism and $g(M)$ is a summand of N, then $H(g) : H(N) \to H(M)$ is surjective.

iii) If R is a field and $g : M \to N$ is a homomorphism, then g is surjective (injective) iff $H(g)$ is injective (surjective).

Proof This is a good exercise.

For the remainder of this section, suppose $W = R_R$. In this case $H(M) = \text{Hom}_R(M, R)$ is denoted by $H(M) = M^*$ and $H(g)$ is denoted by $H(g) = g^*$.

Theorem Suppose M has a finite free basis $\{v_1, ..., v_n\}$. Define $v_i^* \in M^*$ by $v_i^*(v_1 r_1 + \cdots + v_n r_n) = r_i$. Thus $v_i^*(v_j) = \delta_{i,j}$. Then $v_1^*, \ldots, v_n^*$ is a free basis for M^*, called the *dual basis*. Therefore M^* is free and is isomorphic to M.

Proof First consider the case of $R^n = R_{n,1}$, with basis $\{e_1, \ldots, e_n\}$ where $e_i = \begin{pmatrix} 0 \\ \cdot \\ 1_i \\ \cdot \\ 0 \end{pmatrix}$.

We know $(R^n)^* \approx R_{1,n}$, i.e., any homomorphism from R^n to R is given by a $1 \times n$ matrix. Now $R_{1,n}$ is free with dual basis $\{e_1^*, \ldots, e_n^*\}$ where $e_i^* = (0, \ldots, 0, 1_i, 0, \ldots, 0)$. For the general case, let $g : R^n \xrightarrow{\approx} M$ be given by $g(e_i) = v_i$. Then $g^* : M^* \to (R^n)^*$ sends v_i^* to e_i^*. Since g^* is an isomorphism, $\{v_1^*, \ldots, v_n^*\}$ is a basis for M^*.

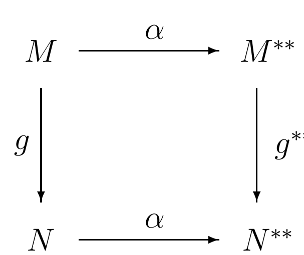
Theorem Suppose M is a free module with a basis $\{v_1, \ldots, v_m\}$ and N is a free module with a basis $\{w_1, \ldots, w_n\}$ and $g : M \to N$ is the homomorphism given by $A = (a_{i,j}) \in R_{n,m}$. This means $g(v_j) = a_{1,j} w_1 + \cdots + a_{n,j} w_n$. Then the matrix of $g^* : N^* \to M^*$ with respect to the dual bases, is given by A^t.

Proof Note that $g^*(w_i^*)$ is a homomorphism from M to R. Evaluation on v_j gives $g^*(w_i^*)(v_j) = (w_i^* \circ g)(v_j) = w_i^*(g(v_j)) = w_i^*(a_{1,j}w_1 + \cdots + a_{n,j}w_n) = a_{i,j}$. Thus $g^*(w_i^*)$ $= a_{i,1}v_1^* + \cdots + a_{i,m}v_m^*$, and thus g^* is represented by A^t.

Exercise If U is an R-module, define $\phi_U : U^* \oplus U \to R$ by $\phi_U(f, u) = f(u)$. Show that ϕ_U is R-bilinear. Suppose $g : M \to N$ is an R-module homomorphism, $f \in N^*$ and $v \in M$. Show that $\phi_N(f, g(v)) = \phi_M(g^*(f), v)$. Now suppose $M = N = R^n$ and $g : R^n \to R^n$ is represented by a matrix $A \in R_n$. Suppose $f \in (R^n)^*$ and $v \in R^n$. Use the theorem above to show that $\phi : (R^n)^* \oplus R^n \to R$ has the property $\phi(f, Av) = \phi(A^t f, v)$. This is with the elements of R^n and $(R^n)^*$ written as column vectors. If the elements of R^n are written as column vectors and the elements of $(R^n)^*$ are written as row vectors, the formula is $\phi(f, Av) = \phi(fA, v)$. Of course this is just the matrix product fAv. Dual spaces are confusing, and this exercise should be worked out completely.

Definition "Double dual" is a "covariant" functor, i.e., if $g : M \to N$ is a homomorphism, then $g^{**} : M^{**} \to N^{**}$. For any module M, define $\alpha : M \to M^{**}$ by $\alpha(m) : M^* \to R$ is the homomorphism which sends $f \in M^*$ to $f(m) \in R$, i.e., $\alpha(m)$ is given by evaluation at m. Note that α is a homomorphism.

Theorem If M and N are R-modules and $g : M \to N$ is a homomorphism, then the following diagram is commutative.



Proof On M, α is given by $\alpha(v) = \phi_M(-, v)$. On N, $\alpha(u) = \phi_N(-, u)$. The proof follows from the equation $\phi_N(f, g(v)) = \phi_M(g^*(f), v)$.

Theorem If M is a free R-module with a finite basis $\{v_1, \ldots, v_n\}$, then $\alpha : M \to M^{**}$ is an isomorphism.

Proof $\{\alpha(v_1), \ldots, \alpha(v_n)\}$ is the dual basis of $\{v_1^*, \ldots, v_n^*\}$, i.e., $\alpha(v_i) = (v_i^*)^*$.

Note Suppose R is a field and C is the category of finitely generated vector spaces over R. In the language of category theory, α is a natural equivalence between the identity functor and the double dual functor.

Note For finitely generated vector spaces, α is used to identify V and V^{**}. Under this identification V^* is the dual of V and V is the dual of V^*. Also, if $\{v_1, \ldots, v_n\}$ is a basis for V and $\{v_i^*, \ldots, v_n^*\}$ its dual basis, then $\{v_1, \ldots, v_n\}$ is the dual basis for $\{v_1^*, \ldots, v_n^*\}$.

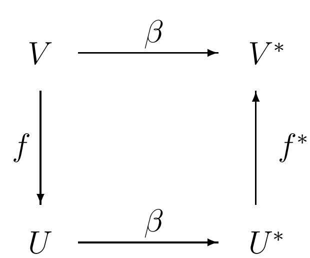
In general there is no natural way to identify V and V^*. However for real inner product spaces there is.

Theorem Let $R = \mathbf{R}$ and V be an n-dimensional real inner product space. Then $\beta : V \to V^*$ given by $\beta(v) = (v, -)$ is an isomorphism.

Proof β is injective and V and V^* have the same dimension.

Note If β is used to identify V with V^*, then $\phi_V : V^* \oplus V \to \mathbf{R}$ is just the dot product $V \oplus V \to \mathbf{R}$.

Note If $\{v_1, \ldots, v_n\}$ is any orthonormal basis for V, $\{\beta(v_1), \ldots, \beta(v_n)\}$ is the dual basis of $\{v_1, \ldots, v_n\}$, that is $\beta(v_i) = v_i^*$. The isomorphism $\beta : V \to V^*$ defines an inner product on V^*, and under this structure, β is an isometry. If $\{v_1, \ldots, v_n\}$ is an orthonormal basis for V, $\{v_1^*, \ldots, v_n^*\}$ is an orthonormal basis for V^*. Also, if U is another n-dimensional IPS and $f : V \to U$ is an isometry, then $f^* : U^* \to V^*$ is an isometry and the following diagram commutes.



Exercise Suppose R is a commutative ring, T is an infinite index set, and for each $t \in T$, $R_t = R$. Show $(\bigoplus_{t \in T} R_t)^*$ is isomorphic to $R^T = \prod_{t \in T} R_t$. Now let $T = \mathbf{Z}^+$, $R = \mathbf{R}$, and $M = \bigoplus_{t \in T} \mathbf{R}_t$. Show M^* is not isomorphic to M.

Index